The Transcriptome of Epithelial Transformation in the Distal Gastrointestinal Tract

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde

(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Jacob Sabatés Bellver

aus

Spanien

Promotionskomitee

Prof. Dr. Josef Jiricny (Leitung der Dissertation)

Prof. Dr. Michael Hengartner

Prof. Dr. Anne Müller

Dr. Giancarlo Marra

Zürich, 2007

Table of contents

1. Summary

Cancer is the second most frequent cause of death in developed countries. The most dangerous ability of cancer cells of a primary tumor is to invade adjacent or distant organs. This process, known as metastasis, represents the main cause of death from cancer. Among all the different cancer types, colorectal cancer is the second cause of death by malignant tumors in both Europe and US.

The carcinogenic process in the colon starts when genetic and/or epigenetic alterations accumulate in normal tissue cells. Some of the mutations affect key genes involved in signal transduction, apoptosis, and DNA repair such as mismatch repair. In most cases, the transformation process, which begins in the epithelial crypts, seems to result from qualitative, quantitative, and spatial subversion of the Wnt signaling pathway, the physiological regulator of epithelial homeostasis. These early events trigger the transformation of normal cells into dysplastic cells, which proliferate and form benign tumors, called **adenomas**. Additional sequentially ordered mutations arise to end up with malignant lesions, named **cancer**. Thus, colon cancer progression can be divided in three critical stages; i) the normal to adenomatous cell transition, ii) the progression from adenoma to carcinoma, and iii) the appearance of metastasis. Clinical intervention, based on the biological knowledge of the different stages acquired during the last three decades, could greatly improve the management of this disease.

In an attempt to discover genes responsible for the initial steps of epithelial transformation, we performed a comprehensive transcriptome analysis of colorectal adenomas (**see Results and Appendix I**).

Among the most over-expressed genes in adenomas, we found *KIAA1199*, a gene with unknown function. In normal tissue, both the RNA and protein were expressed at the bottom of the intestinal crypts where the proliferative compartment is located. Interestingly, we saw a striking over-expression all over the adenomatous tissue, particularly in the cytoplasm facing the crypt lumen and in/around the mucin vacuoles of goblet cells. The over-expression persisted during the transition from low-grade dysplasia with goblet-cell differentiation to high-degree dysplasia in which this differentiation was no longer apparent. KIAA1199 mRNA and protein were also over-expressed in advanced colon cancers.

These data, taken together with other findings obtained during a collaborative work with Prof. Hans Clevers' laboratory, Hubrecht Institute, Netherlands Institute for Developmental Biology, Utrecht (**see Results and Appendix II**), suggested ***KIAA1199*** as a novel biomarker of colorectal tumors, as well as a new target gene of the Wnt signaling pathway.

Briefly, we compared the list of up-regulated genes in adenomas and carcinomas (compared with normal mucosa), with that of genes down-regulated upon stable transfection of LS174T and DLD1 colorectal cancer cell lines with dominant negative TCF1 or TCF4 (dnTCF1 and dnTCF4), or beta-catenin siRNA. These dnTCFs block the formation of a transcription complex composed of beta-catenin and the DNA-binding proteins of the T cell factor/lymphoid-enhancing factor (TCF/LEF) family. We assumed that genes common to both lists might be putative Wnt targets. This combined analysis of microarray data for tissues and cell lines placed *KIAA1199* at the top of the common gene list.

For a better clinical management of tumors, it is essential to know the different molecular characteristics they acquire during transformation. Not all colorectal cancers exhibit the same features in terms of growth rate, aggressiveness, prognosis, etc. For example, defects in one of the DNA mismatch repair genes are crucial events for cancer initiation and progression in the colon and other organs (for details see 3.2.3.). These defects confer characteristic phenotypes to the tumors: high level DNA microsatellite instability, conspicuous lymphocytic infiltration, preferential location in the proximal colon, and better prognosis. It has been generally accepted that the better prognosis of mismatch repair deficient colon cancers is related to the conspicuous infiltration of CD8+ lymphocytes, but the molecular aspects of this histological feature have not been studied. In a microarray analysis of mismatch repair deficient cancers, we have identified ***4-1BBL***, a gene playing a key role in the anti-tumor response, as one of the most prominently over-expressed genes in these tumors. Its over-expression was also confirmed at the protein level with flow cytometry. We have proposed that this ligand presented by neoplastic cells could be a target of 4-1BB present on CD8+ cytotoxic lymphocytes. This finding could also be explored to identify novel approaches for the treatment of this subgroup of colon cancers. We also showed that MSI could conceivably be responsible for the decreased transcript levels of genes containing 5'-UTR or coding- region microsatellites. Alterations in the 5'-UTR could diminish transcription efficiency, whereas coding-region MSI would lead to nonsense-mediated decay of the RNA transcript. In contrast, genes whose 3'-UTRs contained long mononucleotide repeats generally displayed increased expression in MMR-deficient settings, and the 3 genes of this type tested for MSI all displayed repeat shortening. The fact that in reverse-transcription polymerase chain reaction the microsatellite-containing tracts of the 3'-UTRs of these genes (but not their coding regions) were amplified much more efficiently in MMR-deficient cells is difficult to explain, but recent evidence suggests that poly(U)-tract length can affect RNA stability and processing and translation efficiency (**see Results and Appendix III**).

Another aspect that can affect cancer incidence is the intrinsic molecular and biological nature of the normal tissue of origin.

For example, in the lower gastrointestinal tract, concretely in the ileum (the last part of the small intestine) and in the colon (the main part of the large bowel), cancer rates are surprisingly different showing an adenocarcinoma incidence in the ileum around 25 times lower than in the colon, despite the fact that both tissues have high proliferative rates.This striking difference motivated us to perform a comparative transcriptome analysis of these two normal tissues (**see Results and Appendix IV**).

Although this work is still ongoing, preliminary results showed interesting transcriptional differences in i) genes related to cell differentiation and cell proliferation, ii) transcription regulators, iii) Wnt-signaling pathway genes, and iv) apoptosis.

In this last group of genes, two pro-apoptotic genes members of the cell death-inducing DFFA-like effectors family, ***CIDEB*** and ***CIDEC***, were found significantly under-expressed in colon compared with ileum. In addition, transcript levels for these genes were further down-regulated in colon adenomas and adenocarcinomas. These findings raise the question of whether the down-regulation of

pro-apoptotic genes in normal colon might be related to a higher predisposition to cellular transformation in this organ.

In conclusion, during my PhD studies, we have 1) characterized the transcriptome of normal epithelia of the lower gastrointestinal tract, and of benign and malignant tumors of the colon, 2) identified novel molecular markers of adenomatous transformation and of advanced colon cancer with defective mismatch repair system, 3) identified new putative Wnt target genes in colon adenomas and carcinomas.

We strongly believe the data we produced could give important hints to better understand colon carcinogenesis and to improve its clinical treatment.

2. Zusammenfassung

Krebs ist die zweithäufigste Todesursache in entwickelten Ländern. Die wohl gefährlichste Eigenschaft von Krebszellen eines primären Tumors ist es, in umliegende oder entfernte Organe vorzudringen. Dieser Prozess, bekannt unter dem Begriff Metastasieren, ist der Hauptgrund für den tödlichen Verlauf von Krebs. Unter all den verschiedenen Typen ist der kolorektale Krebs die zweithäufigste Todesursache durch bösartige Tumoren, und dies sowohl in Europa als auch in den US.

Durch die Ansammlung von genetischen und/oder epigenetischen Veränderungen in normalen Gewebezellen startet der karzinogene Prozess im Dickdarm. Einige der bekannten Mutationen betreffen Gene, welche Schlüsselfunktionen in Signaltransduktion, Apoptosis und DNA-Reparatur (z.B. Fehlpaarungsreparatur (MMR)) einnehmen. Der Transformationsprozess, welcher im Krypten-Epithel beginnt, scheint in den meisten Fällen das Resultat qualitativer, quantitativer und räumlicher Zerstörung des Wnt-Signalweges zu sein. Der Wnt-Signalweg ist der physiologische Regulator der Epithel-Homöostase. Diese frühen Ereignisse lösen die Transformation von normalen Zellen in fehlgebildete, dysplastische Zellen aus. Letztere proliferieren und bilden gutartige Tumoren, sogenannte **Adenome**. Entstehen weitere sequenziell angeordnete Mutationen führt dies zu bösartigen Läsionen, sprich Krebs. Deshalb kann das Fortschreiten von Dickdarmkrebs in drei kritische Phasen unterteilt werden: i) Transition von normalen Zellen zu adenomatösen Zellen, ii) der Verlauf von Adenoma zu Karzinoma und iii) das Auftreten von Metastasen. Durch das fundierte biologische Verständnis dieser verschiedenen Phasen konnten die klinischen Behandlungsansätze für die Krankheit in den letzten drei Jahrzehnten stark verbessert werden.

In einem Versuch, die für die initialen Schritte der Epithel-Transformation verantwortlichen Gene zu finden, führten wir eine umfassende Transkriptom-Analyse von kolorektalen Adenomen durch (**siehe Results and Appendix I**).

Als eines der am stärksten überexprimierten Gene in Adenomen entdeckten wir *KIAA1199*, ein Gen mit unbekannter Funktion. In normalem Gewebe wird sowohl die RNA als auch das Protein selber zuunterst in den Darm-Krypten exprimiert, dort wo sich das proliferative Kompartement befindet. Interessanterweise sahen wir eine markante Überexprimierung im gesamten adenomatösen Gewebe, speziell in dem dem Krypten-Lumen zugewandten Zytoplasma und in/um die Muzin-Vakuolen der Becherzellen. Die Überexprimierung bestand von der Transition von low-grade Dysplasie mit Becherzell-Differenzierung bis hin zur hochgradigen Dysplasie bei welcher diese Differenzierung nicht mehr vorhanden war. Die mRNA und das Protein von KIAA1199 waren auch in fortgeschrittenem Dickdarmkrebs überexprimiert.

Diese Daten, zusammen mit den Resultaten aus der Kollaboration mit Prof. Hans Clever's Labor, Hubrecht Institute, Netherlands Institute for Developmental Biology, Utrecht (**siehe Results and Appendix II**), lassen den Schluss zu, dass KIAA1199 sowohl ein neuer Biomarker für kolorektale Tumore, als auch ein neues Zielgen des Wnt-Signalweges ist.

Kurz, wir verglichen die Liste der aufregulierten Gene in Adenom und Karzinom (verglichen mit normaler Mukosa) mit der Liste der abregulierten Gene nach stabiler Transfektion von LS174T und

DLD1 kolorektalen Krebszell-Linien mit dominant-negativem TCF1 oder TCF4 (dnTCF1 and dnTCF4), oder beta-catenin siRNA. Die dnTCFs verhindern die Bildung von Transkriptionskomplexen bestehend aus beta-catenin und den DNA-Bindungsproteinen der T cell factor/lymphoid-enhancing factor (TCF/LEF)-Familie. Wir vermuteten, dass Gene, die auf beiden Listen erscheinen, mögliche Wnt-Zielgene sein könnten. Diese kombinierte Microarray-Daten-Analyse für Gewebe und Zell-Linien rückte KIAA1199 an die Spitze der gemeinsamen Gen-Liste.

Um Tumoren in der Klinik besser behandeln zu können ist es unabdingbar, die verschiedenen molekularen Eigenschaften, welche durch Transformation entstehen, genau zu kennen. Nicht jeder kolorektale Krebs zeigt die gleichen Verhaltensweisen in bezug auf Wachstumsgeschwindigkeit, Aggressivität, Prognose etc. Zum Beispiel sind Defekte in einem an der Fehlpaarungs-Reparatur beteiligten Gen eine massgebliche Voraussetzung für den Beginn und das Fortschreiten von Krebs im Dickdarm und anderen Organen (für Details siehe 3.2.3.). Diese Defekte verleihen dem Tumor einen charakteristischen Phänotyp: ein hohes Niveau an DNA Mikrosatelliten-Instabilität (MSI), auffällige lymphozytäre Infiltration, bevorzugte Lokalisation im proximalen Dickdarm und bessere Prognosen. Es ist generell bekannt, dass die bessere Prognose von Fehlpaarungs-Reparatur-defizientem Dickdarmkrebs mit der auffälligen Infiltration von CD8+-Lymphozyten zusammenhängt, wobei die molekularen Aspekte dieser histologischen Eigenart nie studiert wurden. In einer Microarray-Analyse von Fehlpaarungs-Reparatur(MMR)-defizientem Krebs identifizierten wir das Gen ***4-1BBL***, welches nicht nur eine Schlüsselrolle in der Anti-Tumor-Antwort spielt sondern auch eines der am stärksten überexprimierten Gene in diesen Tumoren ist. Die Überexprimierung dieses Genes wurde auch auf Proteinebene mittels Durchfluss-Zytometrie bestätigt. Wir könnten uns vorstellen, dass dieser Ligand von neoplastischen Zellen präsentiert wird und ein Ziel von 4-1BB sein könnte, welches wiederum auf zytotoxischen CD8+-Lymphozyten vorhanden ist. Diese Erkenntnis könnte auch dazu benutzt werden, neue Behandlungsansätze für diese Krebs-Untergruppe zu identifizieren. Wir konnten auch zeigen, dass MSI möglicherweise verantwortlich für die verminderte Transkription von Genen ist, welche Mikrosatelliten entweder im 5'-UTR-Abschnitt oder in den kodierenden Regionen beherbergen. Veränderungen im 5'-UTR-Abschnitt könnte die Transkriptions-Effizienz abschwächen, während Mikrosatelliten in den kodierenden Regionen non-sense mediated decay der RNA-Transkripte zur Folge haben könnte. Auf der anderen Seite zeigten Gene, welche 3'-UTR mit langen Mononukleotid-Wiederholungen enthielten, generell erhöhte Exprimierung in MMR-defizienten Einstellungen. Drei auf MSI untersuchte Gene dieses Typs waren durch Verkürzungen in den Wiederholungen charakterisiert. Die Tatsache, dass in MMR-defizienten Zellen die Mikrosatelliten-enthaltenden Bereiche in den 3'-UTR-Abschnitten dieser Gene wesentlich effizienter mittels reverse-transcription polymerase chain reaction (RT-PCR) amplifiziert wurden, ist schwierig zu erklären. Neueste Erkenntnisse lassen aber vermuten, dass die Länge von poly(U)-Abschnitten die Stabilität, die Prozessierung und die Translations-Effizienz von RNA beeinflussen kann (**siehe Results and Appendix III**).

Die intrinsische molekulare und biologische Beschaffenheit von normalem Ursprungsgewebe ist ein weiterer Aspekt, welcher das Auftreten von Krebs bestimmen kann.

Als Beispiel: Obschon sowohl im Ileum (letzter Teil des Dünndarms, auch Krummdarm genannt) als auch im Dickdarm (auch Grimmdarm genannt) die Gewebe hoch-proliferativ sind, ist das

Auftreten von Adenokarzinomen überraschenderweise im Ileum zirka 25-mal seltener als im Dickdarm. Dieser markante Unterschied hat uns motiviert, eine vergleichende Transkriptom-Analyse von diesen beiden normalen Geweben durchzuführen (**siehe Results and Appendix IV**).

Obschon wir diese Arbeit noch nicht abgeschlossen haben, geben die ersten Resultate interessante Hinweise: wir beobachten transkriptionelle Unterschiede in i) Genen, welche mit Zell-Differenzierung und Zell-Proliferierung in Verbindung stehen, ii) Transkriptions-Regulatoren, iii) Genen des Wnt-Signalweges und iv) Apoptosie.

In der letztgenannten Gruppe von Genen fanden wir zwei pro-apoptotische Mitglieder der cell death-inducing DFFA-like effectors-Familie, nämlich ***CIDEB*** und ***CIDEC***. Verglichen mit dem Ileum waren beide deutlich unterexprimiert im Dickdarm. Dazu kommt, dass bei diesen Genen das Level der Transkripte in Dickdarm-Adenomas und Adenokarzinomen noch weiter abreguliert war. Die Frage stellt sich nun, ob Abregulierung von pro-apoptotischen Genen in normalem Dickdarmgewebe vielleicht mit einer höheren Prädisposition für zelluläre Transformation in diesem Organ verbunden ist.

Zusammenfassend kann man sagen, dass wir während meiner Doktorarbeit 1) das Transkriptom von normalem Epithel des unteren gastrointestinalen Trakts und von gut- und bösartigen Dickdarmtumoren charakterisiert, 2) neue molekulare Marker der adenomatösen Transformation und des fortgeschrittenen Dickdarmkrebs mit defektivem MMR-System identifiziert und 3) ebenso neue mögliche Wnt-Zielgene in Dickdarm-Adenomen und –Karzinomen identifiziert haben.

Wir glauben, dass unsere Daten wichtige Hinweise für ein besseres Verständnis der Dickdarm-Karzinogenese liefern und dadurch die klinischen Behandlungen verbessert werden können.

3. Introduction

According to the World Health Organization (WHO), the top-5 causes of human mortalism in developed countries are i) cardiovascular diseases, ii) cancer (malignant neoplasm), iii) cerebrovascular diseases (stroke), iv) chronic obstructive pulmonary diseases (emphysema, chronic bronchitis), and v) accidents (unintentional injuries).

On the other hand, the situation in developing countries is different because of shorter life span, due to the lower living standards (lack of potable water, food, and access to medicines) and deficient information and education. The shortest life span decreased the incidence of "developed diseases", like cancer, associated with aging. In contrast, HIV-AIDS (on the top of the list) and parasitic diseases (such as diarrhea, malaria), pathologies under control in developed countries, are primary causes of death.

Cancer, being the second most frequent cause of death in developed countries, is a public health problem that costs thousands of billions every year. For this reason, its prevention, detection, treatment, and cure has become a topic of primary interest for thousands of research groups all over the world.

3.1. Cancer

The term cancer encloses a vast group of diseases that can arise in almost any part of the body. The process by which cancer appears and develops is called carcinogenesis and it is characterized by the rapid and uncontrolled proliferation of cells and their capacity to invade adjacent tissues, or to be transported through the blood stream or lymphatic system to occupy distant tissues, in a process known as metastasis. Metastasis is the major cause of death from cancer.

Cancer risk is thought to be related to age because the increased and uncontrolled growth of the cells is caused by mutations in the DNA and DNA damage, more frequent in aging DNA. Although many mutations arise during DNA replication, there are many mechanisms to repair them and also many mutations are necessary for the transformation of a normal cell into a cancer cell. Mutations in the DNA are a result of the interaction between genetic factors of the host and external agents of different origin; physical (ultraviolet and ionizing radiation), chemical (asbestos, tobacco, alcohol…), or biological (virus, bacteria, and parasites) carcinogens. Cancers can be sporadic, most of them, or inherited, which then are called syndromes.

3.1.1. Classification

The abnormal and disorganized growth in a tissue or organ, usually forming a distinct mass of cells, is called neoplasia or tumor. Tumors can be benign or malignant (cancer). Cancers can be classified according to the tissue origin of the tumor: carcinomas, lymphomas and leukemias, sarcomas, mesotheliomas, gliomas, germinomas, and choriocarcinomas.

Carcinomas: these malignant tumors are derived from epithelial cells and are the most common type of cancers. Carcinomas invade surrounding tissues and organs, and may spread to

lymph nodes and distal sites (metastasis). In this group we included breast, prostate, lung and colon cancers. They can be classified by their histopathological appearance:

Adenocarcinomas are malignant tumors originating in the epithelial cells of glandular tissues.

Squamous cell carcinomas show squamous cell differentiation. Squamous cell carcinoma is usually developed in the epithelial layer of the skin and sometimes in various mucous membranes of the body.

Small cell carcinomas are usually associated with the lung, though they can be associated with other topographies, such as in cervical cancer.

Lymphomas and **Leukemias**: these are malignant tumors derived from blood and bone marrow cells characterized by an abnormal proliferation of blood cells, usually white blood cells (leukocytes).

Sarcomas: are cancers that arise from connective tissue, or mesenchymal cells

Mesothelioma: tumors derived from the mesothelial cells lining the peritoneum and the pleura.

Gliomas: are primary central nervous system (CNS) tumors that arise from glial cells. The most common site of involvement of a glioma is the brain, but they can also affect the spinal cord, or any other part of the CNS, such as the optic nerves.

Germinomas: are malignant tumors which most closely resemble germ line cells. These types of neoplasia are usually found in the testes or ovaries. Germinoma found outside these areas may be referred to as dysgerminomas.

Choriocarcinomas: are malignant and aggressive cancers of the placenta. They are characterized by early hematogenous spread to the lungs.

3.1.2. Worldwide incidence and death rates

In Europe and the United States of America (US), cancer incidence and death rates are very similar[1, 2]. Because of the increasing age of the population, the number of cancer cases will continue to increase every year, although the knowledge about these diseases and the mechanisms to prevent, detect, and treat them, have remarkably improved in the last decades.

In Europe, there were almost 3 millions new cancer cases in 2004[1] (Figure 1). The highest incidence percentages were registered for lung (13.2%), colorectal cancer (CRC; 13.0%), and breast (12.8%) cancers. The cancers with the highest mortality were lung (20%), CRC (11.9%), and stomach (8.1%). Dividing the statistics by gender, the most frequent cancers in men affected lung (19.5%), prostate (15.5%), and CRC (12.8%) whereas in women, breast (27.4%), CRC (13.3%), and uterus (9.9%) cancers were the most common malignant tumors observed. In terms of death rates, lung (27.9% and 9.8%, in men and women, respectively), CRC (10.7% and 13.3%, in men and women, respectively), prostate (15.5% in men), and breast (17.4% in women) cancers comprised top-3 for each gender.

Cancer Incidence and Death Rates in Europe (2004)

Incidence		Death	
Lung	13.2%	Lung	20.0%
CRC	13.0%	CRC	11.9%
Breast	12.8%	Stomach	8.1%
Prostate	8.2%	Breast	7.6%
Stomach	5.9%	Prostate	5.0%
Uterus	4.6%	Lymphoma	3.8%
Lymphoma	4.2%	Leukemia	3.1%
Oral Cavity and Pharynx	3.4%	Uterus	2.9%
Leukemia	2.6%	Oral Cavity and Pharynx	2.3%
All other sites	32.1%	All other sites	35.3%
Total Cases: 2886800	**100%**	**Total Deaths: 1711000**	**100%**

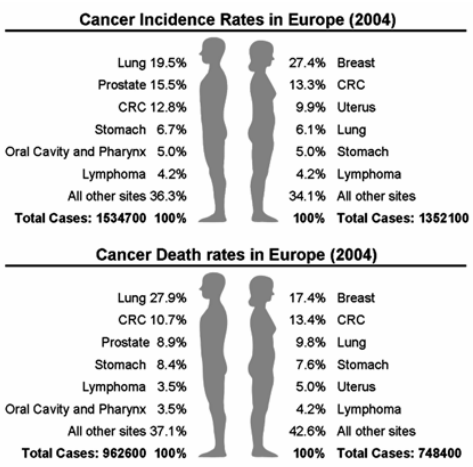


Figure 1. **Cancer's classification** based on incidence and death rates on the **European population** in 2004. Adapted from ref. #1.

In the US, the statistics from 2005[2] were similar to those of the European but with some noteworthy differences (Figure 2). In men, prostate, lung, and CRC were, as in Europe, the three cancer types with the highest number of new cases but while lung and CRC percentages were smaller (13.1% to 19.5% and 10.1% to 12.8%, respectively), the percentage of prostate cancer was the double (32.7% to 15.5%). In women, breast and CRC cancers have similar incidence percentages to the European ones (31.9% to 27.4% and 11.1% to 13.3%, respectively) but lung occurrence is higher (12% to 6.1%). The highest mortality accounts for lung (26.6%), breast (14.7%), and CRC (9.4%), like in Europe, but the deaths due to lung cancer are almost three times more frequent in the female population of US than in the European one.

Cancer Incidence and Death Rates in US (2005)

Incidence		Death	
Prostate	16.9%	Lung	28.7%
Breast	15.4%	CRC	9.6%
Lung	12.6%	Breast	7.1%
CRC	10.6%	Pancreas	5.6%
Urinary Bladder	4.6%	Prostate	5.3%
Melanoma of Skin	4.3%	Leukemia	3.9%
Non-Hodgkin Lymphoma	4.1%	Non-Hodgkin Lymphoma	3.4%
Uterine Corpus	3.0%	Ovary	2.8%
Pancreas	2.3%	Esophagus	1.9%
All other sites	26.2%	All other sites	31.7%
Total Cases: 1372910	**100%**	**Total Deaths: 570280**	**100%**

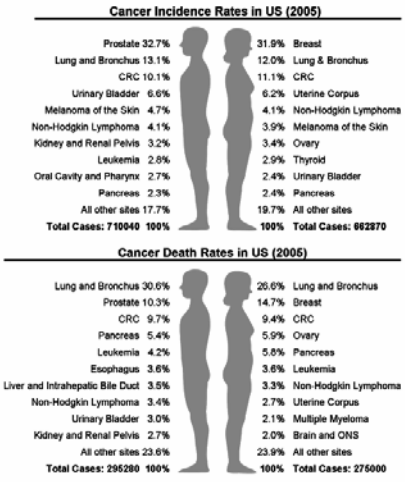


Figure 2. **Cancer's classification** based on incidence and death rates on the **US population** in 2005. Adapted from ref. #2

Interesting information can be extracted from these data. The first striking fact refers to the huge increase of prostate cancer incidence and death rates in the US male population compared to Europe. The second remarkable observation is the enormous increase of the incidence and death rates of lung cancer among US female inhabitants compared to women in Europe. This may be due to

the smoking habits of American women since it has been shown that smoking is the principle and most important risk factor for lung cancer. The third important piece of information points out CRC as the second cause of death by malignant tumors in both Europe and the US. It affects men and women at the same rate (different from lung cancer, for example, that affects many more male than female subjects). Since more than two thirds of the new cases of CRC appear in developed countries, it is believed that this pathology is strongly influenced by social and geographic components. This is one of the reasons why research in the colorectal cancer field is conducted by many laboratories in the world.

3.2. Colorectal Cancer (CRC)

To better understand what is and how CRC arises it is important to have a notion about the physiology and the histology of the intestinal tract (see 3.2.1.). It is also mandatory to have a brief introduction to the intestinal cell renewal mechanism because its (de)regulation is closely linked to cancer onset.

Around 200 cells per crypt (details in 3.2.1.1.) are generated every day in the intestine. This complicated and precise mechanism needs to be strictly controlled to preserve a correct equilibrium. This equilibrium is called homeostasis and it is maintained by, at least, three different coordinated processes: cell proliferation[3], cell migration[4], and anoikis.

The intestinal cell proliferation is not individually driven by each cell type but controlled by the crypt niche, meaning that the two clearly defined proliferative and differentiated compartments are strictly preserved while the cells travel beside the crypt-villi axes[5]. While the location of four to six stem cells[6, 7] at the crypt bottom in the large intestine seems to be clear, their position in the small intestine is less obvious; they have been placed immediately above the Paneth cells[8, 9] or mixed with Paneth cells at the very bottom of the crypts[10-13]. The intestinal stem cells are characterized by preservation of an undifferentiated phenotype, permanent production of all different cell lineages, self-maintenance and renewal abilities, and regeneration capability after damage.

The epithelium layer is subjected to a constant cell migration towards the villi or the surface epithelium. There are only two cell types that escape this effect; Paneth cells and stem cells.

Finally, the processes called anoikis is the cell shedding at the top of the villi, in the small intestine, or at the surface of the epithelium, in the large intestine, by which the cells lose contact with the cellular matrix and undergo apoptosis.

3.2.1. Biology of the gastrointestinal (GI) tract[14-16]

The gastrointestinal tract, also called the digestive tract or gut, is a group of organs that by a procedure called digestion disrupts and processes the food to get energy and nutrients, discarding the remaining waste. In a normal human adult it is approximately 7.5 meters long.

3.2.1.1. Anatomy

The GI tract is composed of two main parts; the upper and the lower gastrointestinal tract, separated, at the same time, in different segments (Figure 3).

The **upper GI tract** encloses the mouth, the esophagus and the stomach.

The **mouth** is the cavity lying at the upper end of the alimentary canal, bounded on the outside by the lips and inside by the oropharynx and containing in higher vertebrates the tongue, gums, and teeth. It is the body opening through which an animal takes in food and water.

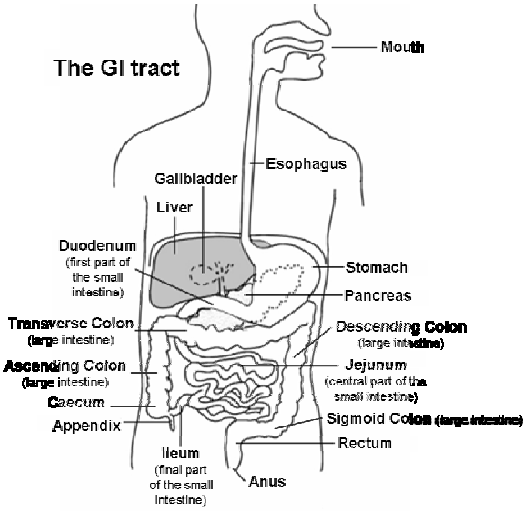


Figure 3. The human gastrointestinal tract.

The human **esophagus** is about 25 cm long and 2.5 cm in diameter. It is the portion of the digestive tube that conducts food from the pharynx to the stomach. When food is swallowed it passes from the pharynx into the esophagus, initiating rhythmic contractions (peristalsis) of the esophageal wall, which propel the food along toward the stomach. The walls of the esophagus are lined with mucous glands that continue the lubrication of the food as it is conducted to the stomach. The **cardia**, or cardiac sphincter, is the anatomical term for the junction orifice of the stomach and the esophagus. At the cardia, the mucosa of the esophagus transitions into gastric mucosa.

The **stomach** is a saclike dilation in the gastrointestinal tract between the esophagus and the intestines, forming an organ of digestion. The human stomach is a muscular, elastic, pear-shaped bag, lying crosswise in the abdominal cavity beneath the diaphragm. It is capable of gross alterations in size and shape, depending on the position of the body and the amount of food inside. The stomach is about 30.5 cm long and is 15.2 cm wide at its widest point. Its capacity is about 0.94 liters in the adult. Food enters the stomach from the esophagus, through the cardiac sphincter that normally prevents food from passing back to the esophagus. The other end of the stomach empties into the first section of the small intestine, or duodenum; the pyloric sphincter, which separates the two, remains closed until the food in the stomach has been modified and is in suitable condition to pass into the small intestine. The muscular action of the stomach and the digestive action of the gastric juice convert food in the stomach into a semiliquid state called chyme. The stomach is involved in the second phase of digestion, following mastication. Its primary function is to break down large molecules (such as from food) into smaller ones using hydrochloric acid so that they can be absorbed by the small intestine. Secondary functions include temporary storage of food, control of the rate at which food enters the duodenum, acid secretion and antibacterial action, fluidization of stomach contents, protein digestion using pepsin enzymes, and absorbing water, some ions, and some lipid soluble compounds such as alcohol, aspirin, and caffeine. The stomach's interior can secrete 2 to 3 liters of gastric fluid per day.

The **lower GI tract**, also called bowel or intestine, is the muscular hoselike portion of the gastrointestinal tract extending from the lower end of the stomach (pylorus) to the anal opening and, in

humans and other mammals, consists of two segments, the small intestine and the large intestine or colon. The large intestine is wider in diameter than the small intestine and its direction as it leaves the cecum is upward (ascending colon), across the abdominal cavity (transverse colon) beneath the stomach, and then downward (descending colon) on the left side of the abdominal cavity, making a sharp turn in the left lower portion (sigmoid colon) to merge with the rectum. Bacteria, the indigestible residue of food, and mucus form the bulk of matter in the large intestine. The water content of the bulk is absorbed through the walls of the large intestine, and the solid matter is excreted through the rectum. The intestines host several kinds of bacteria that deal with molecules which the human body is not able to breakdown itself (Figure 4).

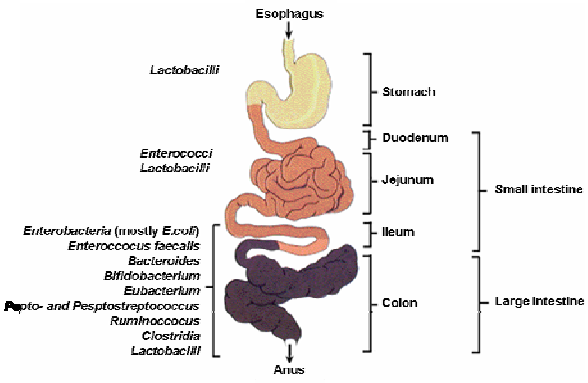


Figure 4. The gastrointestinal flora.

The **small intestine** is the part of the gastrointestinal tract between the stomach and the large intestine. In humans it is about 5-6 m long and 2.5 cm wide. It joins the large intestine (colon) at the cecum in the right lower abdominal cavity. It is divided in three parts; duodenum, jejunum, and ileum. It is covered in wrinkles which are called plicae circulara. From the plicae circulara project microscopic finger-like pieces of tissue called villi. The purpose of these wrinkles and projections is to increase surface area for absorption of nutrients. Furthermore, each villus is covered in microvilli, which again increase the surface area manyfold. In this region, as well, there is the appendix, a blind pouch, of about 7.5 cm long and less than 1.3 cm wide, projecting from the cecum with no digestive function (it is considered to be a remnant of a portion of the digestive tract which was once more functional and is now in the process of evolutionary regression).

The ***duodenum*** is the first and shortest (23–28 cm) segment of the small intestine. It begins with the duodenal bulb and ends at the ligament of Treitz. It curves down and then up from the pylorus of the stomach, where chyme (the thick semifluid mass of partly digested food and secretions, formed in the stomach and intestines during digestion) enters it. Ducts from the pancreas and gallbladder bring in bicarbonate to neutralize stomach acid, pancreatic enzymes to further digestion, and bile salts to break up fats. Nutrient absorption begins in the lower duodenum, which has a mucous lining. Exposure to stomach acid makes the upper duodenum susceptible to peptic ulcers, the duodenum's most common problem.

Following the duodenum there is the ***jejunum***, the central part of the small intestine. In adults it is 1-2 m long and it has a great mobility, like the ileum, thanks to being suspended by the mesentery, a double layer of peritoneum. The inner surface of the jejunum, its mucous membrane, is covered in projections called villi. It differs from the duodenum due to lack of Brunner's glands. It is also different from the ileum due to fewer goblet cells and generally lacks Peyer's patches.

The ***ileum*** is the final and longest (about 2-4 m) segment of the small intestine. It extends from the jejunum to the ileocecal valve, where it joins the cecum. Its function is to absorb vitamin B12

and bile salts. The wall itself is made up of folds, with many villi on its surface. Moreover, the epithelial cells which line these villi possess even larger numbers of microvilli. As a result, the ileum has an extremely large surface area both for the adsorption of enzyme molecules and for the absorption of products of digestion. The cells that line the ileum contain the protease and carbohydrase enzymes responsible for the final stages of protein and carbohydrate digestion. These enzymes are present in the cytoplasm of the epithelial cells. The villi contain large numbers of capillaries which take the amino acids and glucose produced by digestion to the hepatic portal vein and the liver. A difference between the ileum and the other regions of the small intestine is its lymphoid tissues. While the length of the intestinal tract contains lymphoid tissue, only the ileum has abundant Peyer's patches, unencapsulated lymphoid nodules containing large amounts of lymphocytes and other cells of the immune system.

The small intestine has different functions; it is not only an organ of digestion (for that part of the process not completed by the stomach) but is the chief organ of absorption; it is the site where most of the nutrients from ingested food are absorbed. By contraction of its muscular walls (peristalsis) the food mass is propelled onward and, as it is carried along, it is subject to the digestive action of the secretions of the intestinal lining as well as to that of bile and pancreatic juice which enter the upper intestine (duodenum) from ducts leading from the liver and pancreas. The digestion of proteins into peptides and amino acids principally occurs in the stomach but some also occurs in the small intestine. The small intestine is where the majority of chemical digestion takes place; peptides are degraded into amino acids; lipids (fats) are degraded into fatty acids and glycerol; and carbohydrates are degraded into simple sugars (e.g., glucose).

There are many diseases affecting the small intestine. Here, I will just briefly mention two of them: Crohn’s Disease and small intestine cancer.

Cancer of the small intestine is relatively rare compared to colorectal cancer[17]. It can be subdivided into duodenal cancer, and cancer of the jejunum and ileum. The different subtypes of small intestine cancer include adenocarcinomas, gastrointestinal stromal tumor, lymphoma, and carcinoid.

Crohn's disease (CD) is a chronic, episodic disease which can affect any part of the gastrointestinal tract from mouth to anus. Because it is a systemic disease, it can also cause complications outside of the gastrointestinal tract[18]. The main gastrointestinal symptoms are abdominal pain and diarrhea, which may be bloody. Symptoms outside the gastrointestinal tract include skin rashes, arthritis, and ulcers in the mouth. CD is a type of inflammatory bowel disease (IBD). IBD occurs when the immune system contributes to damage of the gastrointestinal tract by causing inflammation. Crohn's disease can be difficult to distinguish from other forms of IBD such as ulcerative colitis. Because of the name, IBD can be confused with irritable bowel syndrome (IBS), a less serious condition. CD affects between 400000 and 600000 people in North America[19]. Prevalence estimates for Northern Europe have ranged from 27–48/100000[20]. Although the cause of Crohn's disease is not known, it is widely believed to be an autoimmune disease. There is a genetic component to susceptibility, and the disease may be triggered in a susceptible person by environmental factors. Unlike the other major type of IBD, ulcerative colitis, there is no known medical

or surgical cure for Crohn's disease. Many medical treatments are however available for Crohn's disease with a goal of keeping the disease in remission[21].

The **large intestine** is the other segment of the lower GI tract. It extends from the end of the ileum to the anus. It is divided into the cecum, the colon, the rectum and the anus. The large intestine is longer in herbivores and shorter in carnivores, and is about 1.2 to 1.8 m long in humans. Its caliber, greater than that of the small intestine, is largest at its commencement at the cecum, and gradually diminishes as far as the rectum, where there is a dilation of considerable size just above the anal canal.

The first part of the large intestine is the ***cecum***. It is a pouch of about 7.5 to 8.5 cm in diameter connected to the ascending colon and the ileum. It is separated from the ileum by the ileocecal valve (ICV) or Bauhin's valve, and is considered to be the beginning of the large intestine.

Nearby the cecum, there is a blind-ended tube called the *appendix*. It develops embryologically from the cecum. It averages 10 cm in length, but can range from 2-20 cm. The diameter of the appendix is usually less than 7-8 mm. While the base of the appendix is at a fairly constant location, the location of the tip of the appendix can vary from being retrocaecal to being in the pelvis to being extraperitoneal. In most people, the appendix is located at the lower right quadrant of the abdomen. Currently, the function of the appendix, if any, remains controversial in the field of human physiology. One explanation has been that the appendix is a remnant of an earlier function, with no current purpose.

The ***colon*** is the part of the intestine located between the cecum and the rectum. It consists of different segments; the ascending colon, the transverse colon, the descending colon and the sigmoid colon, that joins the rectum. The colon from cecum to the mid transverse colon is also known as the right colon. The remainder is known as the left colon.

The first part of the colon is known as the *ascending colon*. The ascending colon expands from the cecum to the hepatic flexure (the turn of the colon by the liver).

Following the hepatic flexure comes the *transverse colon* extending till the splenic flexure (the turn of the colon by the spleen). The transverse colon hangs off the stomach, attached to it by a wide band of tissue called the greater omentum. On the posterior side, the transverse colon is connected to the posterior abdominal wall by a mesentery known as the transverse mesocolon. The transverse colon is encased in peritoneum, and is therefore mobile (unlike the parts of the colon immediately before and after it).

The *descending colon* is the part of the colon from the splenic flexure to the beginning of the sigmoid colon. It is retroperitoneal in two-thirds of humans. In the other third, it has a (usually short) mesentery.

The *sigmoid colon* is the last part of the large intestine before the rectum. Its diameter is approximately 2.5 cm, being the narrowest portion of the colon. The walls of the sigmoid colon are muscular, and contract to increase the pressure inside the colon, causing the stool to move into the rectum.

Next to the sigmoid colon appears the ***rectum***, the final straight portion of the intestine, of about 12 to 15 cm in length, which ends in the ***anus***. The anus is the external opening of the rectum

and is approximately 4 cm long. Closure is controlled by sphincter muscles. Feces are expelled from the body through the anus during the act of defecation, which is its primary function.

The colon has different functions; to lubricate waste products, to absorb remaining fluids (added to the food during digestion), salts and vitamins, and to store waste products until excretion. It has no digestive function (no digestive enzymes are secreted). By the time the chyme has reached the colon, almost all nutrients and 90% of the water have been absorbed by the body in the small intestine. At this point some electrolytes like sodium, magnesium, and chloride are left as well as indigestible carbohydrates known as dietary fiber. As the chyme moves through the large intestine, most of the remaining water is removed, while the chyme is mixed with mucus and bacteria known as gut flora, and becomes feces.

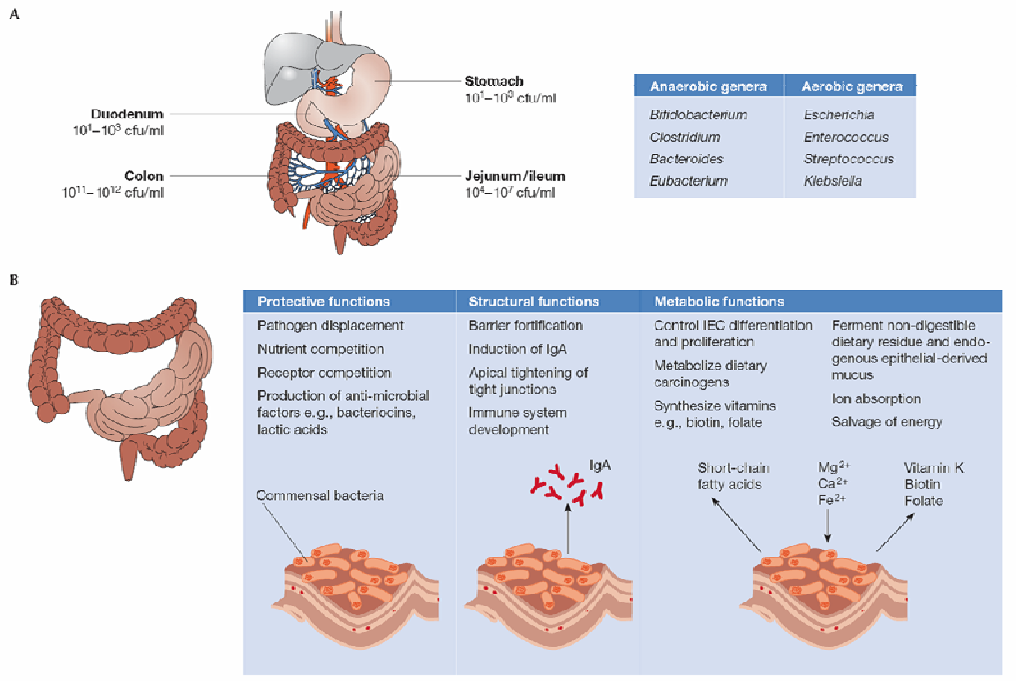


Figure 5. **Functions of the intestinal flora**. (**A**) Bacteria density increases in the jejunum/ileum from the stomach and duodenum, and in the large intestine, colon-residing bacteria achieve the highest cell densities recorded for any ecosystem. The most common anaerobic and aerobic genera are listed. (**B**) Commensal bacteria exert a miscellany of protective, structural and metabolic effects on the intestinal mucosa. Adapted from reference #22.

The bacteria break down some of the fiber for their own nourishment and create acetate, propionate, and butyrate as waste products, which in turn are used by the cell lining of the colon for nourishment. A part of these and other metabolic functions, the gut flora also plays important protective and structural roles[22] (Figure 5).

The colon is affected by several diseases such as Crohn's disease (mentioned above) and cancer (see 3.2.), among many others.

3.2.1.2. Histology

The GI tract has a uniform general histology with some differences which reflect the specialization in functional anatomy. The GI tract can be divided into 4 concentric layers surrounding the inner cavity, the lumen: mucosa, submucosa, muscularis propria (or m. externa), and serosa (Figure 6).

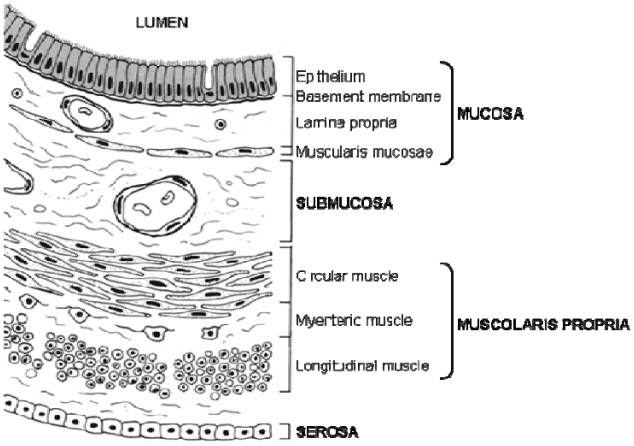


Figure 6. **Histology of the gastrointestinal tract**. Adapted from ref. #15.

The **mucosa** is the innermost layer of the GI tract, surrounding the lumen, or space within the tube. This layer comes in direct contact with the food (or bolus), and is responsible for absorption and secretion, important processes in digestion. The mucosa are highly specialized in each organ of the GI tract, facing a low pH in the stomach, absorbing a multitude of different substances in the small intestine, and also absorbing specific quantities of water in the large intestine. Reflecting the varying needs of these organs, the structure of the mucosa can consist of invaginations of secretory glands, like the gastric pits, or it can be folded in order to increase surface area, for instance villi and microvilli (only present in the small intestine) (Figure 7).

The intestinal villi, fingerlike projections that protrude into the intestinal lumen, are approx. 0.5-1.5 mm long and cover the mucosal surface. The microvilli, sub-light microscopic tubular projections, are extensions of the apical cell membrane and compose the brush border. This complex membranous network contains the enzymes, receptors, and carriers required for terminal digestion and absorption. It can be divided into epithelium, lamina propria (connective tissue core), and muscularis mucosae (layers of smooth muscle).

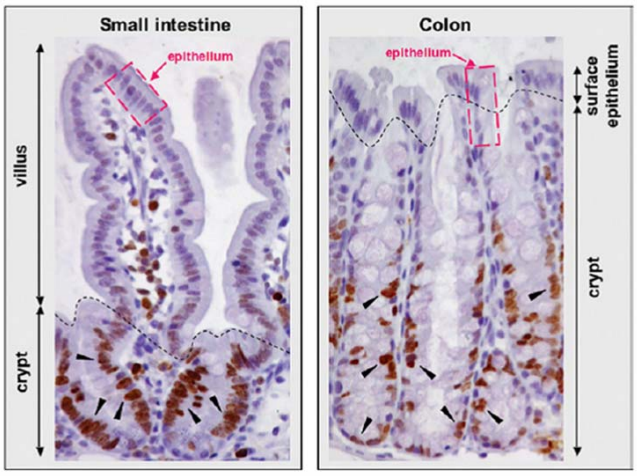


Figure 7. **Morphology of the intestinal and the colonic mucosa of an adult mouse**. The small intestine epithelium is organized into crypts and villi (delimited by dashed lines). In the colon, the crypts are larger, and there are no villi but a flat surface epithelium. The villus and surface epithelium of the colon consist of differentiated epithelial cells, while the crypts are proliferative units as shown by expression of the cell cycle marker KI-67 (arrows). Adapted from reference #40.

The intestines contain a complex, rapidly proliferating, and perpetually differentiating ***epithelium***. The epithelium constitutes the major barrier between the intestinal lumen and the lamina propria and regulates fluxes between these two compartments. Epithelial cell migration and differentiation occurs continuously, and the process of cellular renewal takes approximately three to five days. Cellular differentiation starts during migration to the upper part of the crypt (Lieberkühn crypts) and to the villi base. Microvilli become more prominent, and the cell's capacity to absorb lipids, sugars, and amino acids increases. Shedding cells are removed from the top of the villi (in the small intestine) and the crypts (in the large intestine) by a process known as *anoikis*, in which altered cellular adhesion results in apoptosis (Figure 8).

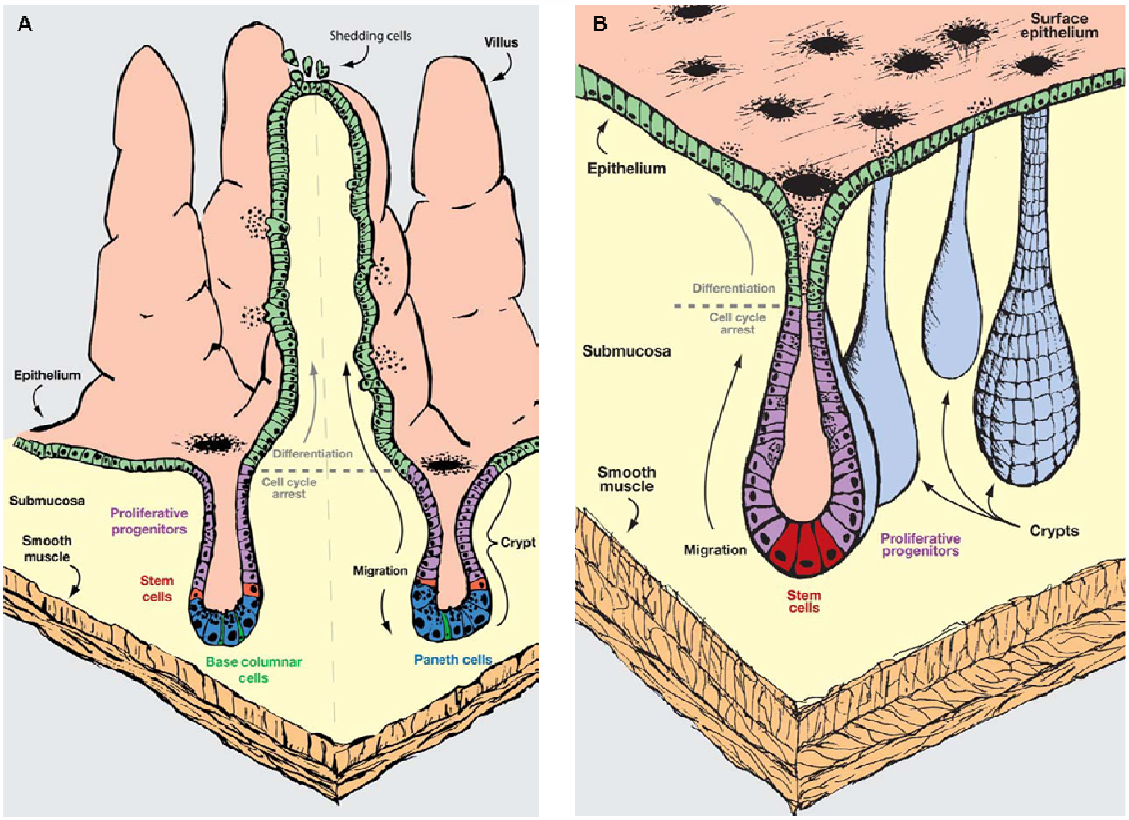


Figure 8. **(A) The structure of the adult small intestine**. Putative stem cells reside immediately above the Paneth cells. Base columnar cells, intermingled between the Paneth cells, may also behave as stem cells. Progenitors stop proliferating at the crypt-villus junction and express differentiation markers. Enteroendocrine, absorptive, and mucosecreting cells migrate upward, whereas Paneth cells migrate downward and localize at the bottom of the crypts. **(B) Structure of the large intestine**. Stem cells reside at the crypt bottom. Progenitors are amplified by constant division along the bottom two thirds of the crypts. Paneth cells are absent in the large intestine. Cell cycle arrest and differentia-tion occur when progenitors reach the top third of the crypts. Adapted from reference #26.

The epithelial cells are replaced by replication of the descendants of epithelial stem cells, believed to reside in the lower crypt region, approximately four to five cell positions above the crypt base (small intestine) or at the bottom of the crypt (large intestine). The anchored stem cells are the source of the four principal differentiated epithelial cell types in the adult intestinal epithelium[23, 24]; the enterocytes (absorptive cells), the goblet cells, the endocrine cells, and the Paneth cells (Figure 9).

The cells that are fated to become the *enterocytes* or *absorptive cells*, the most abundant of the epithelial cells, begin to express a variety of specific genes which enable these cells to digest and absorb many different nutrients.

The role of *goblet cells* is to secrete mucus, a viscous fluid composed primarily of highly glycosylated proteins called mucins suspended in a solution of electrolytes. Mucus serves many functions, including protection against shear stress and chemical damage. They are present throughout the entire GI tract but they are more numerous in the ileum than in the jejunum. These cells exhibit a brandy-goblet shape and are characterized by apically located granules filled with mucins.

The intestine contains remarkably complex types of *endocrine cells*. Although they all arise from a common stem cell, there may be a branch point at which proliferating progenitor cells differentiate into the various endocrine cell lineages. The secretion of these cells may have a multitude of effects on the intestines, like motility and intestinal cellular secretion. They exhibit an appreciable basal surface, and their widths narrow superiorly so that only narrow bands of apical cytoplasm reach the lumen. Their secretory granules are located predominantly in the basal cytoplasm below the nucleus, ready to be secreted by exocytosis through the basal membrane into the lamina propria.

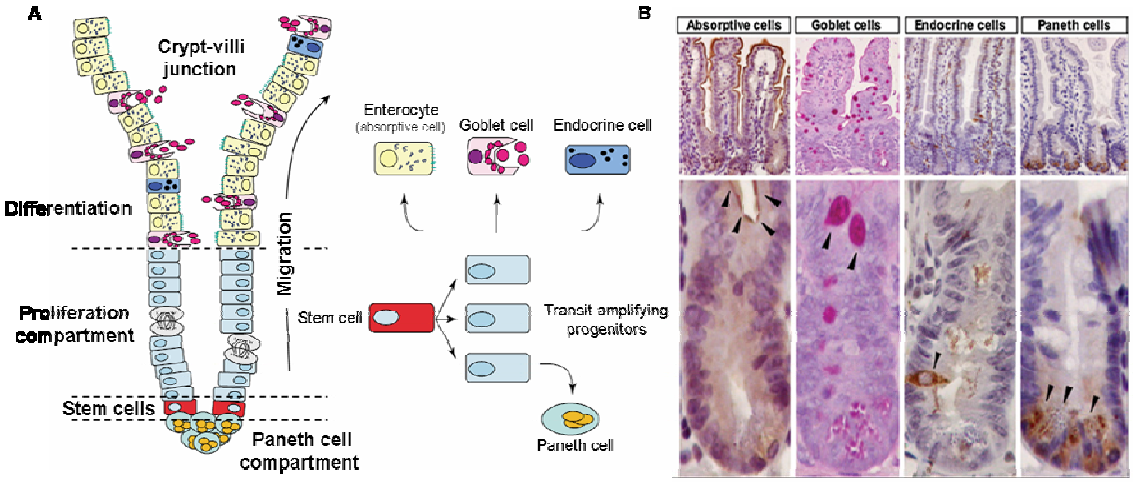


Figure 9. (A) The anatomy of the adult small intestinal epithelium. The epithelium of the adult small intestine is shaped into crypts and villi. The crypt compartment contains the stem cells, the transit amplifying cells and the Paneth cells, whereas the villus compartment is made up of enterocytes, goblet cells and enteroendocrine cells. The pluripotent stem cells give rise to the transit amplifying cells. These cells divide vigorously before they terminally differentiate into one of the four principle cell types: enterocytes, goblet cells, enteroendocrine cells and Paneth cells. **(B) Chemical and immunohistochemical detection of the four principal cell lineages of the small intestine** (arrowheads): villus-associated absorptive cells (alkaline phosphatase), Goblet mucus-secreting cells (periodic acid/Schiff), enteroendocrine cells (synaptophysin), and Paneth cells (lysozyme) (reproduced from Pinto et al., 2003, with permission). The upper small panels show the entire crypt-villus axis while the lower large panels show magnifications of the crypts. Adapted from references #23&24.

Paneth cells are pyramid-shaped cells that reside in the crypt base and contain large eosinophilic secretory granules located in the apical cytoplasm. They are present in the small intestine but not in the colon. Paneth cells provide host defense against microbes in the small intestine. They are functionally similar to neutrophils. When exposed to bacteria or bacterial antigens, Paneth cells secrete a number of antimicrobial molecules into the lumen of the crypt, thereby contributing to maintenance of the gastrointestinal barrier. As said before, small intestinal crypts house stem cells that serve to constantly replenish epithelial cells that die and are lost from the villi. Protection of these stem cells is essential for long-term maintenance of the intestinal epithelium, and the location of Paneth cells adjacent to stem cells suggests that they play a critical role in defending epithelial cell renewal. The principal defense molecules secreted by Paneth cells are alpha-defensins, also known as cryptdin. These peptides have hydrophobic and positively-charged domains that can interact with

phospholipids in cell membranes. This structure allows defensins to insert into membranes, where they interact with one another to form pores that disrupt membrane function, leading to cell lysis. Due to the higher concentration of negatively-charged phospholipids in bacterial than vertebrate cell membranes, defensins preferentially bind to and disrupt bacterial cells, sparing the cells they are functioning to protect. Paneth cells are stimulated to secrete defensins when exposed to bacteria (both Gram positive and negative types) or such bacterial products as lipopolysaccharide, muramyl dipeptide and lipid A. In addition to defensins, Paneth cells secrete lysozyme and phospholipase A2, both of which have clear antimicrobial activity. This battery of secretory molecules gives Paneth cells a potent arsenal against a broad spectrum of agents, including bacteria, fungi and even some enveloped viruses.

The ***lamina propria*** extends from the simple columnar epithelium to the muscularis mucosae and contains many cells involved in immunologic functions. It forms the connective tissue core of the villus and fills the spaces between glands. Primarily a reticular tissue framework with numerous lymphocytes, eosinophils, and plasma cells. It has abundant plasma cells manufacture most of the antibody proteins.

The ***muscularis mucosae*** are several thin layers of smooth muscle fibers located outside the lamina propria and adjacent to the submucosa. They are oriented in different ways which keeps the mucosal surface and underlying glands in a constant state of gentle agitation to expel contents of glandular crypts and enhance contact between epithelium and the contents of the lumen. It aids in the action of continued peristalsis along the gut.

The **submucosa** is a loose connective tissue layer containing numerous arteries as well as venous and lymphatic plexuses that support the mucosa, and muscularis mucosae. It also joins the mucosa to the bulk of underlying smooth muscle. Typical connective tissue cells are also present, but vascular structures predominate, consistent with the role of this portion of the gut as a conduit for absorptive and digested products. Tiny parasympathetic ganglia are scattered around forming the submucosal plexus, or "Meissner's plexus", where preganglionic parasympathetic neurons synapse with postganglionic nerve fibers that supply the muscularis mucosae.

The **muscularis propria** (m. externa) consists of a circular inner muscular layer and a longitudinal outer muscular layer condensed into three bands, the teniae coli. The circular muscle layer prevents the food from going backwards and the longitudinal layer shortens the tract. The coordinated contractions of these layers are called peristalsis and propel the bolus, or balled-up food, through the GI tract. Between the two muscle layers are the myenteric muscle or Auerbach’s plexa.

The **serosa** is a mesothelial cell layer overlying loose connective tissue coated in mucus to prevent friction damage from the intestine rubbing against other tissues. The mesothelium consists of a single layer of vascular flat nucleated cells (simple squamous epithelium) which produce the lubricating serous fluid. This fluid has a consistency similar to thin mucous. These cells are bound tightly to the underlying connective tissue. The connective tissue layer provides the blood vessels and nerves for the overlying secretory cells, and also serves as the binding layer which allows the whole serous membrane to adhere to organs and other structures.

The four concentric layers mentioned above, surround the interior part of the GI tract, called lumen. The **lumen** is the cavity where digested material passes through and from where nutrients are absorbed. The luminal content of the gut is mostly water and therefore chemically represents an aqueous phase. Depending on the location of the gut, the aqueous phase contains acid (i.e., stomach), digestive enzymes (i.e., mouth, esophagus, stomach, and small intestine), or bacteria (i.e., predominatly large intestine).

3.2.2. Signaling Pathways involved in CRC

Proliferation, cell fate specification, differentiation, migration, and cell death in the intestinal tract are precise and complex processes that are regulated through highly synchronized and interconnected developmental programs[25, 26].

The evolutionary conserved Wnt[27, 28], Notch[29, 30], TGF-Beta[31, 32], BMP[33, 34], Hedgehog[35, 36], and RAS/RAF[37, 38] signaling pathways have been shown to be key players in the maintenance and control of homeostasis and anoikis of the small and large intestines.

Many sporadic tumors as well as hereditary syndromes arise and progress due to deregulation (i.e., activation or inactivation) of these signaling cascades. This fact points out how essential and crucial is the correct and controlled functionality of these pathways is in distinguishing the thin line that separates the accurate proliferation/differentiation equilibrium from the apparition of tumor pathologies in these tissues.

Wnt signaling pathway

The Wnt signaling pathway is highly conserved among all metazoa. Our knowledge about this pathway comes from genetic approaches done in nematode, Drosophila, Xenopus, zebrafish, and mouse[39], combined with biochemical approaches (Figure 10).

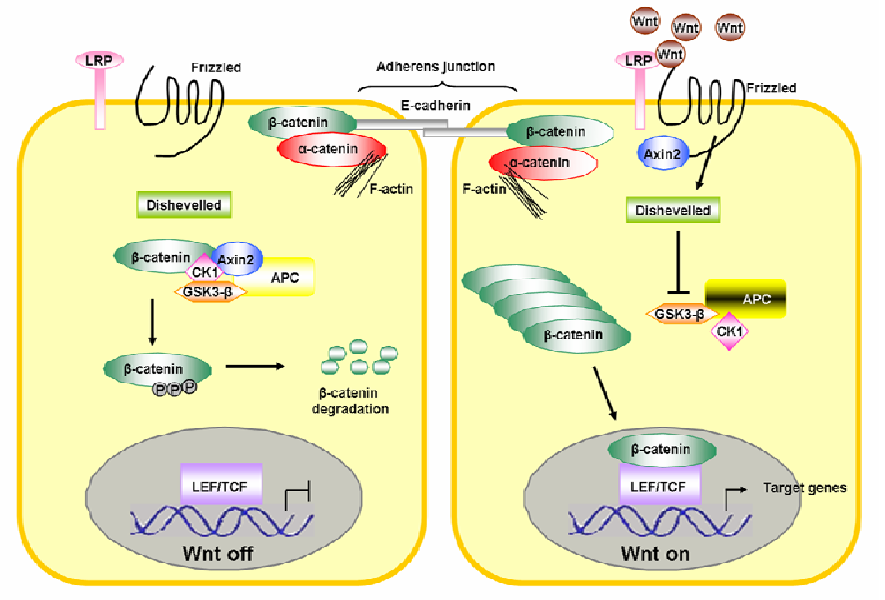


Figure 10. **The Wnt signaling pathway**.

The Wnt signaling pathway is the physiological regulator of epithelium homeostasis[24, 40, 41]. The canonical Wnt signaling pathway is activated when the secreted Wnt glycoproteins bind to the

complex formed by the frizzled (Fz; seven transmembrane receptors) and LRP5/6 (single transmembrane co-receptors members of the low-density lipoprotein receptor related family) receptors.

There are secreted frizzled-related (SFRPs) proteins that bind to the Wnt ligands and block the interaction with Fz proteins. LRP5/6 can also be blocked by other factors (for example, Dkk-1). When the Wnt ligands can not interact with the Fz and LRP5/6 receptors, APC, together with AXIN2, CK1, GSK3-beta, and others, forms a complex in the cytoplasm. CK1 and GSK3-beta sequentially phosphorilate beta-catenin, the central player of this cascade, which is targeted for degradation by the ubiquitin-proteosome pathway.

When the signaling cascade is activated by the Wnt ligands the destruction complex is not active anymore leading to beta-catenin accumulation. Then, beta-catenin translocates into the nucleus, where it interacts with the DNA-binding proteins of the T cell factor/lymphoid-enhancing factor (TCF/LEF) family to form a transcription complex involved in transcriptional activation of many target genes.

Several sources of evidence suggest that Wnt signaling to be activated in intestinal proliferative cells[42-44]; i) nuclear accumulation of beta-catenin, ii) presence of CRC after mutational activation of the signaling cascade, iii) loss of proliferation after mutation of the TCF/LEF family member TCF4[45], iv) change of CRC cells phenotype (from a crypt-like phenotype to a differentiated villi epithelial phenotype) after inhibition of beta-catenin/TCF4 activity[46, 47], and v) reduction of epithelial proliferation and loss of crypts after targeted expression of Wnt inhibitors[48].

The link between Wnt signaling deregulation and CRC has been largely investigated. It has been shown that mutations in APC gene resemble in upregulation of beta-catenin nuclear and cytoplasmic levels[49, 50]. Mutations in AXIN2 or beta-catenin itself also lead to accumulation of beta-catenin in the nucleus and activation of the bet-catenin/TCF4 transcription complex[41, 51]. Many studies have been done to find out Wnt target genes but the final list is still far from completion[47, 52].

Many animal models have been created to try to better understand this pathway and its implication in intestinal cancer, the *Min* mouse model (Min: multiple intestinal neoplasia; is a mutant allele of murine *APC* encoding a nonsense mutation in codon 850) being the most commonly used. Another mutation, the $APC^{\Delta 716}$, gives a truncated protein of 716 amino acids. Heterozygous mice for the *Min* or the $APC^{\Delta 716}$ mutations get dozens of adenomas[53-55].

Notch signaling pathway

Intestinal homeostasis is under the control of various developmental signaling pathways one of which is the Notch signaling pathway. Many cell fate decisions and differentiation processes are regulated by *Notch* genes.

Notch genes are single transmembrane receptors conserved through evolution. There are four different Notch receptors (NOTCH1/2/3/4) and five

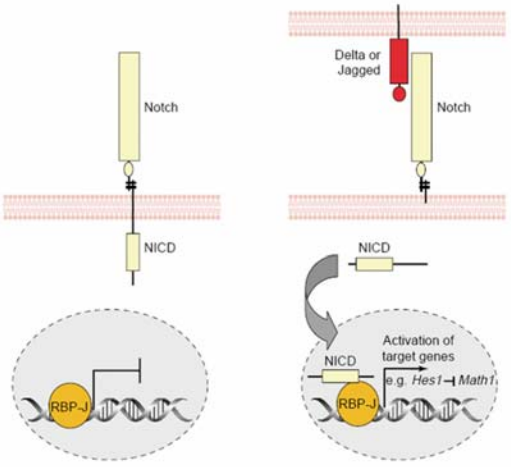


Figure 11. **The Notch signaling pathway**. Adapted from reference # 23.

ligands (DLL1/3/4; JAG1/2)[56]. When receptors are activated by transmembrane ligands of neighboring cells [29], a cascade of proteolytic cleavages of the receptor close to and within the transmembrane domain starts. This leads to the release and translocation of the Notch intracellular domain (NICD) into the nucleus where the transcription factor CSL (CBF1/RBPjk) engages and activates transcription[56, 57] (Figure 11).

Lack of Notch signaling makes CSL to function as a transcriptional repressor. In clusters of precursor cells, Notch controls cell fate choices between adjacent cells. The signal cascade is triggered when a cell reaches higher levels of ligands than the surrounding neighboring cells. There are many known Notch target genes but some of the best characterized ones are members of the hair/enhancer of split (HES) family of transcriptional repressors, which are nuclear basic helix-loop-helix (bHLH) proteins. In turn, these HES proteins regulate downstream genes[58, 59].

Much work has been done with animal models to understand the Notch signaling and its implications in intestinal cell fate[60-68].

TGF-Beta signaling pathway

Many important biological processes, such as angiogenesis, cell differentiation and proliferation, wound healing, and embryonic development, are controlled and regulated by the TGF-beta signaling pathway[32, 69, 70].

The family of TGF-beta cytokines is vast and includes the TGF-beta family members, bone morphogenic proteins (BMPs), and activins. The two surface serine-threonine kinase receptors (type I and type II) are brought together by the ligands and this leads to the phosphorylation of the first receptor by the second. There are three different classes of SMAD proteins, which are the intracellular messengers of this pathway: receptor-regulated SMADs (R-SMADs: SMAD1/2/3/5/8), common SMAD (co-SMAD: SMAD4), and inhibitory SMADs (I-SMADS: SMAD6/7). The association between activated type I receptors and SMAD4 cause R-SMADs phosphorylation and translocation into the nucleus (Figure 12). There, transcription of target genes is regulated by the interaction between the SMAD complex with coactivators or corepressors of transcription. SMAD2/3 are the primary mediators in TGF-beta signaling whereas SMAD1/5/8 mediate BMP signaling.

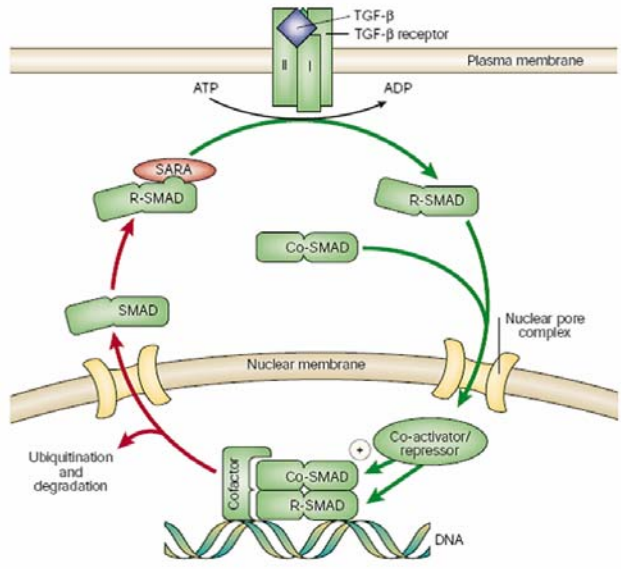


Figure 12. **The TGF-beta signaling pathway**. Adapted from reference #70.

TGF-beta signaling has been shown to be altered in CRC, concretely at the adenoma/carcinoma transition[71]. The inactivation of the TGF-beta receptor type II (TGFBR2) is the most frequent mutation and affects either microsatellite stable (MMS) and unstable (MMI) tumors[72], although SMAD2 and SMAD4 are also mutated in CRC[73, 74]. A number of animal models have been

created to investigate the function of TGF-beta signaling in development and disease. Inactivating mutations for TGFB1/2/3, TGFBR2, SMAD2, SMAD3, and SMAD4 have been reported[75, 76]. Interestingly, in a SMAD4 deficient homozygosity, benign adenomas in APC$^{\Delta 716}$ mice progressed to carcinomas very quickly[77]. Based on the results obtained in animal models, the absence of TGF-beta signaling more than to initiate tumorigenesis, accelerates the development of already existant early tumors.

BMP signaling pathway

As well as the above mentioned TGF-beta signaling pathway, BMP modifies target gene transcription by receptor–mediated intracellular signaling[34] (Figure 13).

There are two types of BMP receptors; type I (BMPR1A and BMPR1B) and type II (BMPR2). In the canonical BMP pathway, BMP receptors type I activate SMAD1/5/8 by phosphorylation[78]. A dimer of phosphorylated R-SMADs then forms a complex with SMAD4. This heterodimic complex translocates into the nucleus and binds to transcription factors regulating the expression of target genes[31, 69]. A non-canonical BMP pathway is driven by TGFh1 activated tyrosine kinase 1 (TAK1, a MAPKKK) and through mitogen activated protein kinase (MAPK)[31, 79]. In addition, X-linked inhibitor of apoptosis (XIAP) links the BMP receptor signal to TAK1, and TAK1 binding proteins (TAB1/2/3) are also required for BMP-mediated TAK1 activation[80, 81]. TAK1 can also trigger Jun N-terminal kinase (JNK) and NF-kB[82].

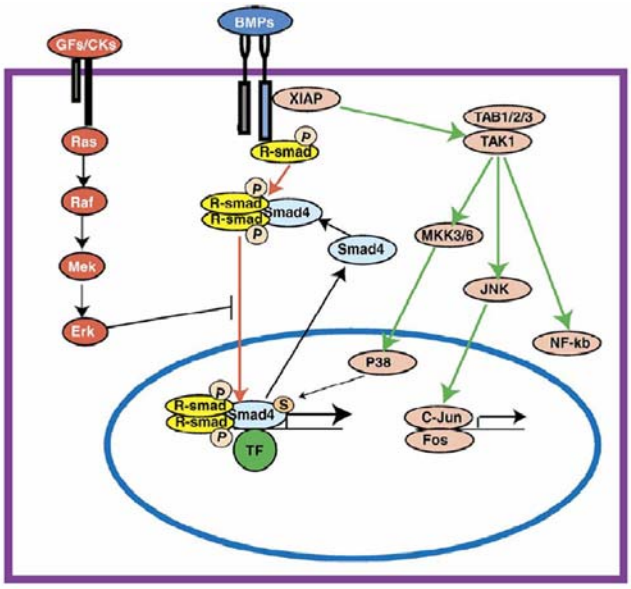


Figure 13. **The BMP signal transduction pathway**. Adapted from reference #34.

In the canonical BMP pathway, SMAD function can be inhibited through blocking its translocation to the nucleus by Erk, in response to GF/CK signaling through Ras/Raf/Mek[83]. Germline mutations affecting SMAD4 or BMPR1A are associated to 50% of Juvenile Polyposis Syndrome (JPS)[84-86].

In this case animal models also help to understand the mechanism of this pathway and its role in JPS[87].

Hedgehog signaling pathway

Hedgehog ligands control cell fate specification during development.

The 12-transmembrane protein called Patched (PTCH) is the receptor for secreted Hedgehog proteins (HH). In the absence of ligand, PTCH inhibits Smoothened (SMO), a 7-transmembrane protein, blocking the signaling cascade[88]. When HH binds to PTCH[89] releases SMO inhibition, leading to activation of the GLI family of Zn-finger transcription factors (GLI1/2/3)[90] (Figure 14). Activated GLI accumulates in the nucleus[91] and controls the transcription of Hedgehog target genes[92-94].

Mutations in the Hedgehog pathway are associated with multiple malformations along the GI tract[95-97]. The development of sporadic tumors of the skin, cerebellum, and skeletal muscle are related to activating mutations of the Hedgehog signaling pathway[98, 99].

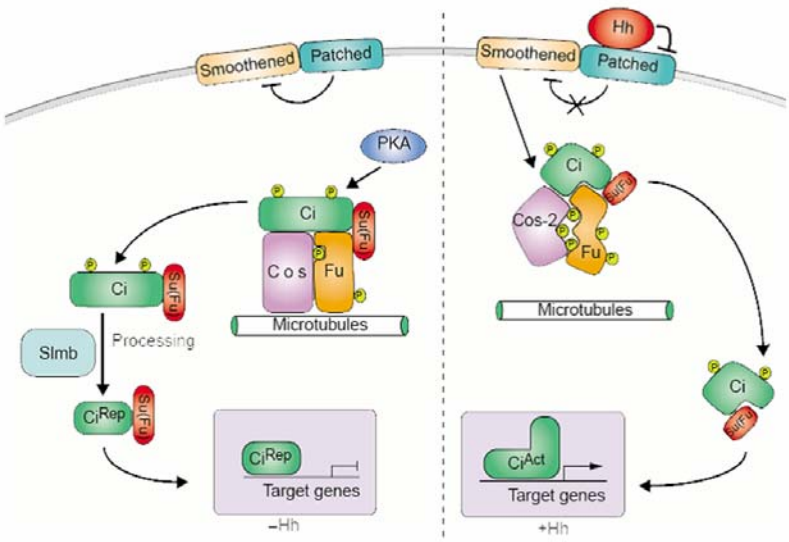


Figure 14. **The Hedgehog signaling pathway**. Adapted from reference #88.

The autocrine/paracrine activation of the Hedgehog pathway in the esophagus, stomach, biliary tract, or pancreas (all of them members of the upper GI tract) seems to be critical for tumor growth[100, 101].

RAS/RAF signaling pathway

The activation of the RAS/RAF/MAPK cascade supports cell proliferation and cell differentiation. RAS genes are GDP/GTP-regulated binary switches.

The formation of the active GTP-bound form of RAS and its conversion to the inactive GDP-bound state is mediated by guanine exchange factors (GEFs) and GTPase activating proteins (GAPs), respectively. Active GTP-bound RAS interacts with RAF (RAF1, A-RAF, or B-RAF) to form a high-affinity complex. As a result of this interaction, RAF gets activated (Figure 15). Then a number of consecutive phosphorylation events take place; RAF phosphorylates MEKs (MEK1/2), which sequentially activates MAPKs (p42 and p44, also called extracellular regulated kinases; ERKs) by phosphorylation. Once phosphorylated, MAPKs translocate into the nucleus to phosphorylate a number of transcription factors, triggering gene expression.

Many mutations in K-RAS and B-RAF have been associated with CRC. Around 50% of CRC carry an activating mutation in K-RAS[102, 103]. Complementation between K-RAS and B-RAF mutations has been described (20% of the CRC with no K-RAS mutation have B-RAF activated[104, 105]) and suggests a key role for the RAS/RAF/MAPK cascade in colorectal tumorigenesis. Interestingly, K-RAS mutations are more frequent in MMR-proficient tumors, whereas B-RAF mutations occur mostly in MMR-deficient lesions[105].

Mutations in K-RAS and B-RAF seem to be related more to cancer progression than to cancer initialization because they are not present in small adenomas[102, 105]. The function of the

RAS/RAF/MAPK pathway in cell renewal of the intestinal epithelium seems to be related to transduction of signals from the membrane to the nucleus.

Work with animal models has not elucidated a key role for K-RAS in the initial stages of cancer in the intestine[106-110].

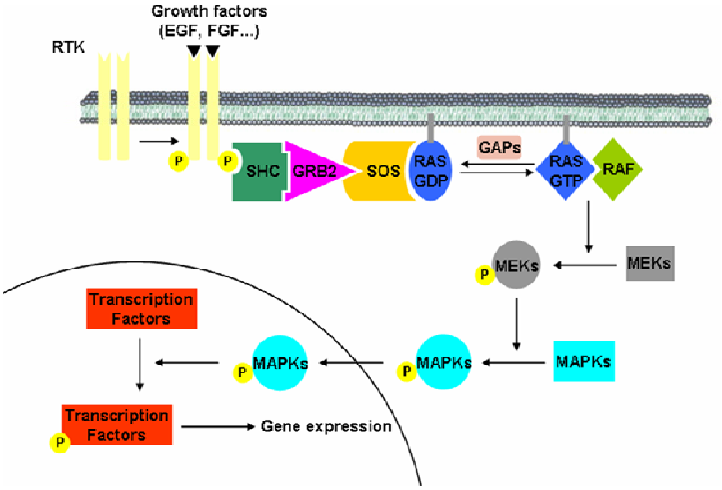


Figure 15. **The RAS/RAF signaling pathway**.

After this overview of the pathways that control and regulate physiologic homeostasis of the intestines, there are evidences that show that cancer cells have developed abilities to mutate key genes on these pathways to assure cancer progression. Furthermore, crosstalk mechanisms between them facilitate their activation through different ways, making more difficult the finding of drugs that could effectively block their signaling cascades to prevent cancer formation.

3.2.3. DNA repair mechanisms, Mismatch Repair (MMR), and MMR defects and Cancer

The processes by which damage present in DNA molecules is detected and eliminated are known as DNA repair mechanisms. Every day, a cell must deal with a continuous onslaught of DNA lesions. This damage can come from environmental or endogenous sources. Both DNA damage sources result in between $1x10^3$ and $1x10^6$ individual molecular lesions per cell per day[111].

The vast majority of DNA damage implies a chemical modification of the bases, the primary structure of the double helix. A failure in repairing these kinds of lesions can trigger a harmful mutation rate, genomic instability, or cell death. Damage in DNA repair genes or in genes responsible of cell cycle regulation can lead to aggressive and invasive diseases (e.g., cancer). DNA damage during replication before cell division can provoke base misincorporation and therefore unrecoverable inheritance of mutations by the daughter cells (unless gene conversion, a rare case of back mutation, occurs).

DNA repair rate depends on many factors, including cell type, age of the cell, and extracellular environment. Those cells carrying large amounts of DNA damage, or those in which the DNA repair system is not able to repair the incoming DNA damage, can follow three different processes: i)

senescence (or irreversible state of dormancy), ii) apoptosis (or programmed cell death), or iii) cancer (i.e., deregulated cell division and proliferation).

As said, there are two different sources of DNA damage; endogenous DNA damage (i.e., damage generated by the byproducts of cellular metabolism: oxidation of bases and generation of strand interruptions by reactive oxygen species, alkylation of bases by endogenous alkylating agents, hydrolysis of bases such as deamination, depurination and depyrimidation, and DNA single- and double-strand breaks resulting from collapsed DNA replication forks or from oxidative destruction of deoxyribose residues[112, 113]), and exogenous DNA damage (such as chemical agents, UV radiation, and ionizing radiation).

Depending on the type of damage done to the DNA's double helical structure, a variety of repair strategies have evolved to restore lost information[114]. If possible, cells use the unmodified complementary strand of the DNA or the sister chromatid as a template to losslessly recover the original information. Without access to a template, cells use an error-prone recovery mechanism known as translesion synthesis as a last resort.

Some types of DNA damage are so common that they have their own cellular subsystem dedicated to counteracting them. These mechanisms do not require a template, since the types of damage they counteract can only occur in one of the four bases. Such direct reversal mechanisms are specific to the type of damage incurred.

The damage can affect only one or both strands of the double helix. When damage consists on a break in both strands there are two mechanisms to repair it; non-homologous end-joining (NHEJ), and homologous recombination (HR) repair[114, 115]. When only one of the two strands is carries a defect, the other strand is used as a template for repairing. The mechanisms that do this kind of repair are base excision repair (BER), nucleotide excision repair (NER), and mismatch repair (MMR)[114, 115]. Because our institute is mainly focused on the study of the MMR system, I will explain in more detail how this mechanism works in prokaryotic and eukaryotic organisms.

3.2.3.1. *E.coli* and eukaryotic MMR mechanisms

The major function of the postreplicative MMR system is to eliminate biosynthetic errors (base-base mispairs and short insertion/deletion loops, IDLs) from newly synthesized DNA. If not repaired, these errors will lead to base-substitutions and frameshift mutations.

The term “postreplicative” implicates MMR in the correction of errors coming from DNA replication. This is due to the fact that the fidelity of polymerases implicated in DNA replication is high (incorrect nucleotide incorporation rate around $1x10^{-4}$-$1x10^{-6}$ before intrinsic exonucleolytic proofreading activity and $1x10^{-7}$-$1x10^{-8}$ after it[116]) but not enough to produce a copy of the entire genomic DNA without errors. MMR ensures error free duplication of the human genome by reducing the error rate to $1x10^{-9}$-$1x10^{-10}$.

Even though great advances in the biochemical and structural aspects of MMR have been achieved in the last years, the precise biological functions of the key players of the human MMR system have not yet been elucidated.

MMR proteins are divided in two categories; MutS homologues (MSH: hMSH2, hMSH3, and hMSH6), and MutL homologues (MLH: hMLH1 and hPMS2, where PMS stands for Post Meiotic Segregation). Other MLH/PMS homologues are also known (hMSH4, hMSH5, hPMS1, and hMLH3) but their roles are largely uncharacterized. MutS and MutL homologues need to form complexes to be functional (Table 1). MMR proteins, DNA replication factors (replication protein A (RPA), proliferating cell nuclear antigen (PCNA), replication factor C (RFC), DNA polymerase delta: Polδ), and exonucleolytic enzymes (exonuclease-1: EXO1) functionally interact in the human MMR machinery.

Table 1. **Human MutS and MutL homologue complexes that are involved in mismatch repair**. Adapted from reference #117.

Complex	Components	Function
MutSα	MSH2, MSH6	Recognition of base–base mismatches and small IDLs
MutSβ	MSH2, MSH3	Recognition of IDLs
MutLα	MLH1, PMS2	Forms a ternary complex with mismatch DNA and MutSα; increases discrimination between heteroduplexes and homoduplexes; also functions in meiotic recombination
MutLβ	MLH1, PMS1	Unknown
MutLγ	MLH1, MLH3	Primary function in meiotic recombination; backup for MutLα in the repair of base–base mismatches and small IDLs

IDL, insertion/deletion loop; MLH, MutL homologue; MSH, MutS homologue; PMS, post-meiotic segregation protein.

The MMR is a highly conserved system from bacteria to human. Such a level of evolutionary conservation is due to its pivotal role in the maintenance of genomic stability. For this reason, much of our understanding of mammalian MMR came from studies done on *E.coli* and *S.cerevisiae* (revised in[117-119]).

MMR in *E.coli*

There are two most important conditions the MMR system has to fulfill; efficient recognition of base-base mismatches and IDLs, and application of the repair machinery to the newly synthesized DNA strand, where the error has been incorporated. How these two criteria are accomplished has been reveled in *in vitro* studies using the *mutS*, *mutL*, *mutH*, and *uvrD E.coli* mutated strains[120].

When a mismatch occurs, a MutS homodimer, via ATP hydrolysis, recruits a homodimer of MutL to activate MutH, which then binds to a hemi-methylated GATC site (it is normally methylated on adenine, but transiently unmethylated because deoxyadenine methylase lags behind the replication fork by approximately 2 minutes) (Figure 16A).

Once the unmethylated strand has been cleaved by the Muts/MutL-depending activated form of MutH, the unwinding of the ends of the error-containing strand from the template by UvrD helicase allows one of the 5'-3' (RecJ or ExoVII) or 3'-5' (ExoI, ExoVII or ExoX) exonucleases, depending on where the nearest hemi-methylated GATC site lies from the mismatch, to degrade the unwound DNA.

After the mismatch has been removed, the resynthesis is mediated by DNA polymerase III and DNA ligase.

Eukaryotic MMR mechanism

Even though the MMR system is conserved, the mechanism in eukaryotes is more complex than in *E.coli*. Five human MutS homologues (hMSH2, hMSH3, hMSH4, hMSH5, and hMSH6) have

been identified although only hMSH2, hMSH3, and hMSH6 take part in MMR (hMSH4, and hMSH5 are exclusively involved in meiotic recombination).

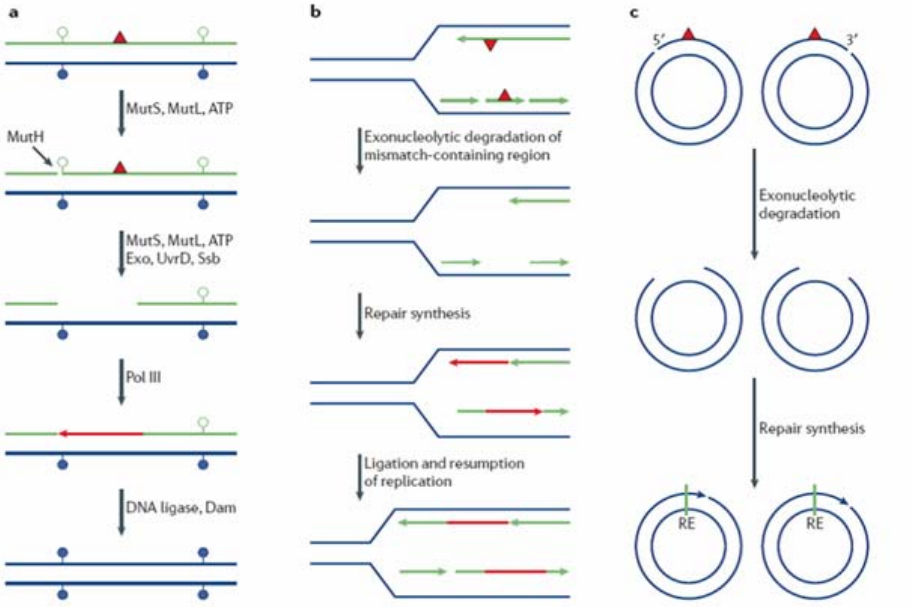


Figure 16. **Postreplicative mismatch repair**. **a** | In *Escherichia coli*, newly synthesized DNA (green) is transiently unmethylated at GATC sites (empty circles). The mismatch (red triangle)-activated MutS–MutL–ATP complex licenses the MutH endonuclease to incise the nearest unmethylated GATC sequence (either 5' or 3' from the mismatch). UvrD helicase, together with one of several exonucleases (Exo), generate a gap that extends from the nick to ~100 nucleotides past the mispair, and which is stabilized by the single-stranded DNA-binding protein Ssb. This gap is filled (shown in red) by DNA polymerase III (pol III) and the remaining nick is sealed by DNA ligase. The GATC sites are then methylated (solid blue circles) by deoxyadenine methylase (Dam). **b** | In mammalian cells, the degradation of the mismatch-containing leading strand (green) might begin at the 3' terminus of the primer strand. Once the mispair is removed, the polymerase resynthesizes the degraded region. Mismatch repair (MMR) in the lagging strand might remove an entire Okazaki fragment, with degradation commencing at either end. Extension of the fragment closest to the replication fork would replace the degraded one. Removal of the RNA termini of the Okazaki fragments, followed by ligation, would give rise to a continuous, error-free, lagging strand. **c** | Substrates used in the *in vitro* MMR assays14,15. These circular heteroduplexes contain a single-strand break that is either 5' or 3' from a mismatch, which lies within a recognition sequence of a restriction enzyme (RE), and therefore makes the molecule refractory to cleavage by this enzyme. Mismatch-provoked degradation of the discontinuous strand, followed by repair synthesis, restores the RE-cleavage site. Adapted from reference #117.

Likewise, four MutL homologues are known in human cells (hMLH1, hMLH3, hPMS1, and hPMS2). Both types of homologues participate in the MMR in the form of heterodimers[118]. The different complexes and their known functions are described in Table 1.

A particularly interesting characteristic of hMutL and hMutS complexes is that hMLH1 and hMSH2 are stable by themselves, whereas hPMS1, hPMS2, and hMSH6 need the presence of the corresponding interacting partner to be stable (in absence of hMLH1: hPMS1, hPMS2; and in the absence of hMSH2: hMSH6, degrade). hMutLα, formed by hMLH1 and hPMS2, has the most important role in MMR (reviewed in[121]) and has been the most studied heterodimer. The role of hMutLγ in MMR has also recently been investigated[122-124].

As mentioned above, an important criterion that the MMR system has to satisfy is to direct the repair machinery to the mismatch-carrying newly synthesized strand. Interestingly, and different than in *E.coli* where MutH provides strand specificity by recognizing hemi-methylated GATC sequences, no MutH functional homologue has been found in other organisms. This fact may imply that the processing of mismatches that occur during replication could be triggered by strand breaks such as

the 5' or 3' termini of Okazaki fragments in the lagging strand, or the 3' terminus of the leading strand (Figure 16B).

The statement that something similar happens with mismatches arising during recombination, where the repair could be directed to the invading strand by its 3' terminus (Figure 16C) is supported by *in vivo*[125, 126] and *in vitro*[120] experimental work with MutH deficient *E.coli* and by human *in vitro* MMR assays[127-129].

Based on the biochemical knowledge acquired so far, three different mammalian MMR models have been proposed: i) the molecular-switch model[130], favors the stochastic bidirectional diffusion of multiple MutSα and MutLα sliding clamps from the mismatch. The repair process would start when a strand break tagged with other MMR features is encountered by clamps, ii) the active-translocation model[131, 132], which hypothesizes that MutSα, probably with MutLα, after being released from the mismatch, translocates, in a controlled ATP hydrolysis-dependent manner, along the DNA, iii) the DNA bending/verification model, claims that MSH proteins remain in the proximity of the mismatch and, rather than protein movement along the DNA, is communication between the mismatch and the strand discrimination signal which involves DNA bending, although experimental evidences showed no satisfactory conclusions about that model[133, 134]. Since the first two models appear to be the most reliable ones, and because they only differ in the way MutSα is translated from the mismatch, they have been represented as one in Figure 17 (see Figure 17 legend for details).

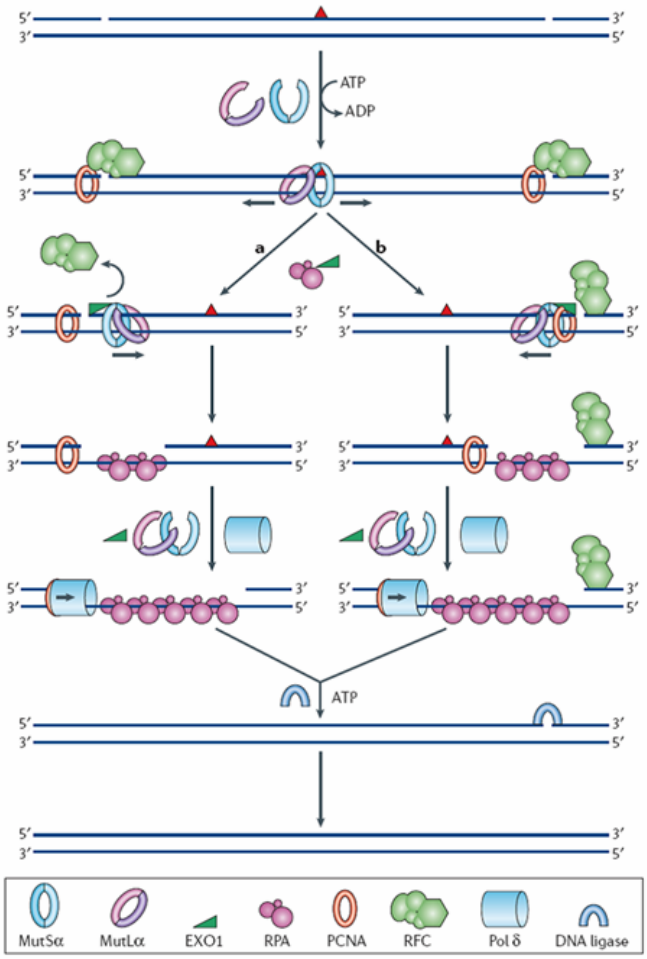


Figure17. **The reconstituted human mismatch-repair system.** The mismatch repair (MMR) process was recently reconstituted72,74 from either MutSα or MutSβ, MutLα, replication protein A (RPA), exonuclease-1 (EXO1), proliferating cell nuclear antigen (PCNA), replication factor C (RFC), DNA polymerase δ (Pol δ) and DNA ligase I. The following is proposed to take place. The mismatch (red triangle)-bound MutSα (or MutSβ) recruits MutLα. The ternary complex undergoes an ATP-driven conformational switch, which releases the sliding clamp from the mismatch site. **a** | Clamps that diffuse upstream encounter RFC that is bound at the 5' terminus of the strand break, and will displace it and load EXO1. The activated exonuclease commences the degradation of the strand in a 5'→3' direction. The single-stranded gap is stabilized by RPA. When the mismatch is removed, EXO1 activity is no longer stimulated by MutSα, and is actively inhibited by MutLα. Pol δ loads at the 3' terminus of the original discontinuity, which carries a bound PCNA molecule. This complex fills the gap and DNA ligase I seals the remaining nick to complete the repair process. **b** | Clamps that migrate downstream encounter a PCNA molecule that is bound at the 3' terminus of the strand break. The recruitment and the activation of EXO1 results in the degradation of the region between the original discontinuity and the mismatch, possibly through several iterative EXO1-loading events. RFC that is bound at the 5' terminus of the discontinuity prevents degradation in the 5'→3' direction (away from the mismatch). Once the mismatch is removed and the EXO1 activity is inhibited by bound RPA and MutLα, the gap is filled by Pol δ. DNA ligase I seals the remaining nick to complete the repair process. Adapted from reference #117.

3.2.3.2. MMR defects and cancer

There are two different classes of mutations derived from the inactivation of the MMR system and the consequent decrease in the repair rates of the total number of errors after DNA replication: simple nucleotide misincorporation, and strand misalignments.

Nucleotide misincorporations take place when DNA polymerase incorporates an erroneous base creating a mismatch (for example, incorporation of thymine opposite to guanine giving as a result a G/T mismatch) which escapes 3'-5' exonuclease proofreading activity. When the MMR system is not mutated, it recognizes the mismatch and repairs it. In cells with a deficient MMR system, the mismatch is not recognized resulting in a G:C to T:A transversion mutation incorporated in half of the progeny DNA.

Strand misalignments are due to slippage of the DNA polymerase complex during replication. This phenomenon is predominantly frequent in microsatellite sequences (i.e., short repetitive sequences) and generates "insertion/deletion" loops which are recognized and repaired by MMR. If not repaired, such misalignments lead to frameshift mutations. Based on experimental work with MMR-deficient cells, the most recurrent mutation is the loss of one unit in mononucleotide and dinucleotide repeats. Such an event is known as microsatellite instability (MSI)[135, 136], and represents a hallmark of MMR-deficient tumors.

To give uniformity to MSI analysis, a panel of five microsatellite markers, two mononucleotide repeats (BAT25 and BAT26) and three dinucleotides repeats (D2S123, D5S346, and D17S250), known as Bethesda markers, are most commonly used[137]. According to these markers, the degree of microsatellite instability of a tumor can be high (MSI-H), when two or more markers show instability, or low (MSI-L), when microsatellite is detected in fewer markers. Cells carrying defects in MMR genes have increased the risk of malignant transformation.

Several anticarcinogenic functions of MMR genes are compromised in MMR-deficient cells: First, decreased genome stability, manifested as MSI all over the genome[138, 139]. Second, loss of heterology-dependent suppression of recombination, a fact that implies gene conversion and exposure of tumor suppressor genes to loss heterozygosity[140], or chromosomal translocations[141]. Third, increased mutation rates in critical "target genes" (i.e., those with short microsatellite sequences in their codong sequences) involved in growth suppression[142], apoptosis[143], signal transduction[144, 145], etc., resulting in growth advantage for the cells. MMR defects have been related to both hereditary and sporadic colon cancers (discussed in 3.2.4. and 3.2.5.), as well as to sporadic cancers in other organs called "sporadic tumors of the HNPCC spectrum" and "sporadic tumors not belonging to the HNPCC spectrum", respectively[146].

3.2.4. Hereditary Syndromes

The familial form of colorectal cancer has become a major public health problem due to its relatively high frequency among western population. From the total number of CRC cases, ~80% are sporadic (see 3.2.5.), and ~20% are familial or inherited forms[147].

Among the familial cases, i) 10-15% are probably caused by gene mutations and polymorfisms of low penetrance, of which the I1307K polymorphism in the APC gene is the best

example, ii) ~1% are tumors with chromosomal instability (CIN), with a tendency to appear on the left side colon, showing aneuploid DNA, carrying characteristic mutations (in *APC*, *KRAS*, and *p53*), and behaving aggressively, being familial adenomatous poliposis (FAP) the most common example, and iii) between 5% and 8% are tumors classified as HNPCC based on the Amsterdam and Bethesda criteria, (Figure 18).

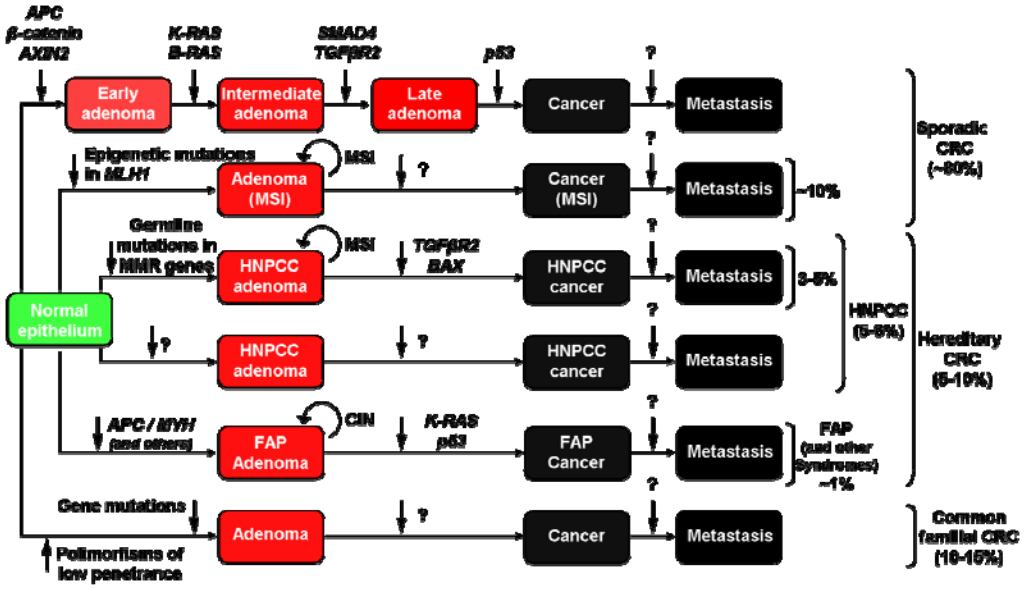


Figure 18. The molecular mechanisms of sporadic and hereditary colorectal carcinogenesis.

Among the HNPCCs, accounting for 3-5% of total CRC, there are cases with germline mutations in one allele of a MMR gene (predominantly in *hMLH1* and *hMSH2*, but also in *hMSH6*, and *hPMS2* genes). These cancers show MSI, arise predominantly on the right side colon, with diploid DNA, harboring characteristic mutations (in *TGFBR2*, and *BAX*), and behaving less aggressively. The remaining HNPCC and HNPCC-like cases are still molecularly unexplained. Several important studies have contributed to the increase in our knowledge about these hereditary syndromes and many reviews have put together what is known thus far[44, 71, 102, 121, 137, 146-161].

Below, FAP will be briefly described and HNPCC will be more extensively discussed.

In **FAP** patients, numerous (hundreds, even thousands) adenomas arise, mainly in the epithelium of the colon, at an early age. If not removed, the initial benign lesions undergo malignant degeneration by the age of 40 to 50 years in most of the patients (~100%).

FAP is due to germline mutations in the *APC* gene. Although the polyps are inherently benign, the first step of the two-hit hypothesis[162] has already taken place: the inherited APC mutation. Often, the remaining wild type (wt) allele is mutated or deleted, accelerating the generation of polyps. Further mutations (e.g. in *KRAS* or *p53*) to APC-mutated cells are much more likely to lead to cancer than they would in non-mutated epithelial cells.

When FAP results from mutations in the *APC* gene, it is inherited in an autosomal dominant (AD) pattern, which means one copy of the altered gene is sufficient to cause the disorder. In most cases, an affected person has one parent with the condition.

HNPCC

Hereditary non-polyposis colorectal cancer syndrome, formerly known as Lynch syndrome, is the most common form of inherited predisposition to colon cancer.

It is characterized by the development of colorectal cancers but also by the apparition of endometrium cancer, and, less frequently, stomach, small intestine, urinary tract (kidney, renal pelvis, and ureter), ovary, and brain cancers[147].

Table 2. **Amsterdam I and Amsterdam II Criteria**. Adapted from reference #144.

Amsterdam I criteria
At least three relatives must have histologically verified colorectal cancer: One must be a first-degree relative of the other two. At least two successive generations must be affected. At least one of the relatives with colorectal cancer must have received the diagnosis before the age of 50 years. Familial adenomatous polyposis must have been excluded.
Amsterdam II criteria
At least three relatives must have a cancer associated with hereditary non-polyposis colorectal cancer (colorectal, endometrial, stomach, ovary, ureter or renal-pelvis, brain, small-bowel, hepatobiliary tract, or skin [sebaceous tumors]): One must be a first-degree relative of the other two. At least two successive generations must be affected. At least one of the relatives with cancer associated with hereditary non-polyposis colorectal cancer should have received the diagnosis before the age of 50 years. Familial adenomatous polyposis should have been excluded in any relative with colorectal cancer. Tumors should be verified whenever possible.

The early age of onset (between the age of 40 and 50) and the high penetrance (~80%), due to an autosomal dominant transmission trait, are the clinical hallmarks of this syndrome. Based on that, Amsterdam criteria I[160] and II[161] (Table 2), and Bethesda guidelines[163, 164] (Table 3) are the international diagnostic criteria for HNPCC. The Bethesda guidelines are more sensitive but less specific than the Amsterdam criteria in identifying HNPCC families with pathogenic mutations[165, 166].

Another characteristic of HNPCC cases is acceleration in the carcinogenic process. In the general population, the adenoma/carcinoma transition takes 8 to 10 years, whereas an adenoma may emerge as a carcinoma in only 2 to 3 years in the HNPCC subjects[151, 167].

Table 3. **Classical and revised Bethesda criteria**. Adapted from reference #156.

Classical Bethesda criteria
1. Individuals with two HNPCC-related cancers, including synchronous and metachronous colorectal cancers or associated extra-colonic cancers (endometrial, ovarian, gastric, hepatobiliary, small bowel cancer or transitional cell carcinoma of the renal pelvis or ureter).
2. Individuals with colorectal cancer and a first degree relative with colorectal cancer and/or HNPCC-related extra-colonic cancer and/or colorectal adenoma; one of the cancers diagnosed at age < 45 years, and the adenoma diagnosed at age < 40 years.
3. Individuals with colorectal cancer or endometrial cancer diagnosed at age < 45 years.
4. Individuals with right-sided colorectal cancer with an undifferentiated pattern on histopathology diagnosed at age < 45 years.
5. Individuals with signet-ring-cell-type colorectal cancer diagnosed at age < 45 years.
6. Individuals with adenomas diagnosed at age < 45 years.
Revised Bethesda criteria
1. Colorectal cancer diagnosed in a patient < 50 years of age.
2. Presence of synchronous, metachronous colorectal, or other HNPCC-associated tumors[a], regardless of age.
3. Colorectal cancer with MSI-H histology diagnosed in a patient < 60 years of age.
4. Individual with colorectal cancer and one or more first-degree relatives with an HNPCC-related tumor, with one of the cancers diagnosed under age of < 50 years.
5. Individual with colorectal cancer and two or more first-degree or second-degree relatives with HNPCC-related tumors, regardless of age

[a]HNPCC-related tumors include colorectal, endometrial, stomach, ovarian, pancreas, ureter, renal pelvis, biliary tract, and brain tumors, sebaceous gland adenomas and keratoacanthomas in Muir–Torre syndrome and carcinoma of the small bowel

As mentioned before, a substantial portion of the HNPCC cases carry a germline mutation in one allele of a MMR gene (*hMLH1*, *hMSH2*, *hMSH6*, or *hPMS2* genes) whereas the molecular mechanism responsible for the remaining HNPCC and HNPCC-like cancers is still unknown. I will focus my attention on the first group.

Such mutations are detected in ~80% of HNPCC families[168-170], primarily affecting *hMLH1* (~50%), *hMSH2* (~40%), and *hMSH6* (~10%), and less frequently *hPMS2*[171] and *hMLH3*[172] (Table 4).

Table 4. **Genes associated with HNPCC predisposition by virtue of hereditable mutations found in those genes.** Adapted from reference #143.

Gene	Phenotypic Features Associated With Germline Mutations
MLH1	Mostly typical HNPCC. ~30% of mutations are of the missense type whose phenotypic manifestations may vary.
MSH2	Mostly typical HNPCC. Extracolonic cancers may be more common than in MLH1 mutation carriers. A major gene underlying Muir-Torre syndrome.
MSH6	Typical or atypical HNPCC. Often characterized by late onset, frequent occurrence of endometrial cancer, distal location of colorectal cancers, and low degree of MSI in tumors.
PMS2	Typical or atypical HNPCC. The penetrance of mutations may vary. A major gene underlying Turcot's syndrome.
MLH3	Mostly atypical HNPCC. May be characterized by distal location of colorectal cancers and variable degree of MSI in tumors.
EXO1	Mostly atypical HNPCC. May be associated with MSI in tumors.

Abbreviations: HNPCC, hereditary nonpolyposis colon cancer; MSI, microsatellite instability.

These mutations, called "first hit", are not enough to cause MSI, but a "second hit", or somatic mutation, which inactivates the wt allele of the gene, turns on the mutator phenotype (the mutator phenotype is a hypothesis that postulates that the mutation rate in the early stages of tumor development must be significantly greater than that of normal somatic cells) and the subsequent MSI. If the mutation is located in an intergenic region of the gene it may have no biological relevance, but if it is sited in a coding region causing a frameshift mutation, biological consequences emerge.

Inactivation of genes involved in signal transduction (*TGFBR2*)[142], and apoptosis (*BAX*)[143] by deletions in poly(A) repeats of their codon regions, favors cell transformation rather than cell death, therefore cancer progression. Other genes have been reported to be altered in lower frequency (IGFR2[173], TCF4[174], AXIN2[145], GARE[175], and B2M[176]).

Although these lesions have aggressive histological characteristics (such as poor cell differentiation), their prognosis is quite favorable due to lower rates of metastatic potential[146, 177, 178]. This fact could be explained by i) the high number of cytotoxic T-lymphocytes present in these tumors, which, activated by the mutated peptides of the cancer cell surface, trigger tumor cell apoptosis, and ii) a lower MMR-deficient cell viability due to spontaneous mutagenesis, although the molecular mechanisms of this event are still unclear.

The evolution of the two inherited CRC syndromes discussed above, shows interesting variations. In FAP, thousands of benign tumors form after *APC* mutation. Each of them slowly progresses to malignancy, requiring the sequential accumulation of mutations in *KRAS* and *p53*, etc. Among so many adenomas, some of them will progress to cancer. When not treated, the median age of cancer diagnosis is 42 (the same as in HNPCC), 25 earlier than the median age of sporadic colorectal cancer patients.

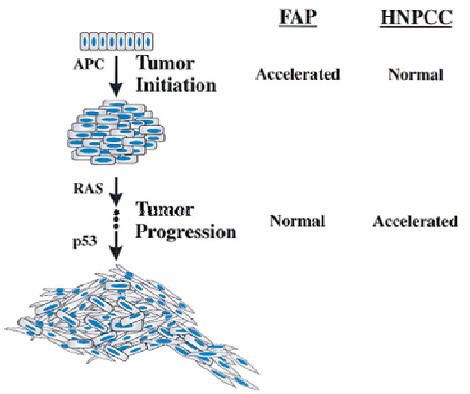


Figure 19. Comparison of the Development of Cancer in FAP and HNPCC Patients FAP results from an increased rate of tumor initiation due to abrogation of the gatekeeper function of APC. In contrast, the mismatch repair defect in HNPCC results in an enhanced rate of mutation that greatly accelerates tumor progression but results in near normal rates of tumor initiation. That the rate of tumor initiation is nearly normal in HNPCC patients is indicated by the fact that such patients do not have a greatly increased incidence of adenomas, in striking contrast to FAP patients (Lynch et al., 1996). But FAP and HNPCC patients both develop colorectal cancers at a median age of 42 years, suggesting that tumor progression is as rate-limiting for cancer formation as is tumor initiation. Adapted from reference #44.

In HNPCC, adenomas form at approximately the same rate as in the general population. Nevertheless, the adenomatous cells with the MMR defect gain mutations at a rate two of three orders of magnitude higher than in normal cells. The accumulation of mutations in tumor suppressors and oncogenes leads to a rapid progression to cancer. Therefore, FAP can be considered as a disease of tumor initiation and HNPCC one of tumor progression (Figure 19).

3.2.5. Sporadic CRC

Around 10% of all sporadic colorectal cancers show MSI. Differently from HNPCC tumors, in which MMR genes are mainly inactivated by mutations, MSI in sporadic CRC is nearly totally due to epigenetic inactivation of *hMLH1* expression by promoter hypermethylation[179].

Experimental work with cell lines showed that the hypermethylation affects, and thus inactivates, both alleles at the same time[180]. The fact that this phenomenon is already present in non-neoplastic colorectal mucosa and colorectal adenomas, suggests that it is an early event in the process of cancer growth[179, 181, 182].

As mentioned before, CRC can be defined as MSI-H (high degree of MSI), MSI-L (low degree of MSI), and MSS (microsatellite-stable tumors). The group of MSI-H colorectal cancers include tumors that; i) arise predominantly on the proximal colon, ii) have diploid DNA content, iii) are poorly differentiated, and iv) have better survival[183-186]. The clinicopathological characteristics do not seem to be able to distinguish MSI-L from MSS tumors[187-189], but there is not an agreement about whether MSI-L tumors should be considered separate from MSS tumors or not.

On the other hand, it has been shown that methylation inactivates other genes besides *hMLH1*.

Hence, tumors in sites other than the colon can appear through different tumorigenic pathways. Despite this fact, the histology and biology of such tumors are very similar, thus, they are identified as "sporadic tumors of the HNPCC spectrum" (Table 5).

Table 5. **Involment of the MMR genes in sporadic tumors of the HNPCC spectrum**. Adapted from reference #143.

Investigators	Origin of Cancer	MMR Gene Studied	Proportion (%) of Tumors With Alterations
Kuismanen et al	Colorectum	MLH1	36/46 (78%)
		MSH2	7/46 (15%)
Cunningham et al	Colorectum	MLH1	48/51 (94%)
		MSH2	3/51 (6%)
		MSH6	4/51 (8%)
Simpkins et al	Endometrium	MLH1	12/14 (86%)
		MSH2	2/14 (14%)
Berends et al	Endometrium	MLH1	3/12 (25%)
		MSH2	3/12 (25%)
Baek et al	Stomach cancer	MLH1	14/16 (88%)
		MSH2	0/16 (0%)
	Gastric adenoma	MLH1	13/15 (87%)
		MSH2	1/15 (7%)
Leung et al	Stomach	MLH1	10/11 (91%)
		MSH2	0/11 (0%)
Chiaravalli et al	Ovary	MLH1	3/3 (100%)
		MSH2	0/3 (0%)

Table 6. **Involment of the MMR genes in sporadic tumors not belonging to the HNPCC spectrum**. Adapted from reference #143.

Investigators	Origin of Cancer	MMR Gene Studied	Proportion (%) of Tumors With Alterations
Benachenhou et al	Breast	MLH1	10/22 (46%)
		MSH2, MSH6	1/20 (5%)
		MSH3	5/22 (23%)
		PMS1	2/20 (10%)
		PMS2	1/21 (5%)
Chen et al	Prostate	MLH1	Low
		MSH2	Low
		MSH6	Low
		PMS1	7/14 (50%)
		PMS2	Moderate
Yeh et al	Prostate	MLH1	Moderate
		MSH2	Moderate
		MSH6	Low
		MSH3	Low
		PMS1	Low
		PMS2	Low
Xinarianos et al	Lung	MLH1	88/150 (59%)
		MSH2	85/147 (58%)
Thykjaer et al	Bladder	MLH1, MSH2, MSH6, MSH3, PMS1, PMS2	4/12 (33%)
Kassem et al	Bladder/TCC	MLH1	1/24 (4%)
		MSH2	1/24 (4%)
	Bladder/SCC	MLH1	0/12 (0%)
		MSH2	1/12 (8%)
Uchida et al	Esophagus	MLH1, MSH2, MSH6, MSH3, PMS2	3/22 (14%)
Wang et al	Head and neck (SCC)	MLH1, MSH2	33/57 (58%)

A considerable portion (2% to 50%[137, 146, 178]) of other sporadic cancers (breast, prostate, lung, bladder, and others), known as “sporadic tumors not belonging to the HNPCC spectrum”, also carry MMR defects (Table 6). Among these tumors, and different from HNPCC and sporadic tumors of the HNPCC spectrum, MMR defects do not seem to start the malignant process because they arise in late stages of tumorigenesis. Nevertheless, MSI status can still be used as a genetic prognostic marker in these cancers to choose the optimal therapeutic strategies for each case.

However, around 90% of the sporadic colorectal cancers do not carry MSI. The sequential model for CRC genesis[71] (Figure 20) claims that mutations in the APC gene, and less frequently in the beta-catenin-encoding gene and AXIN2, are required for tumor initiation. Subsequent progression towards malignancy is accompanied by other mutational alterations, affecting oncogenes (KRAS), tumor suppressor genes (p53), and other unknown genes, as well as epigenetic events[190].

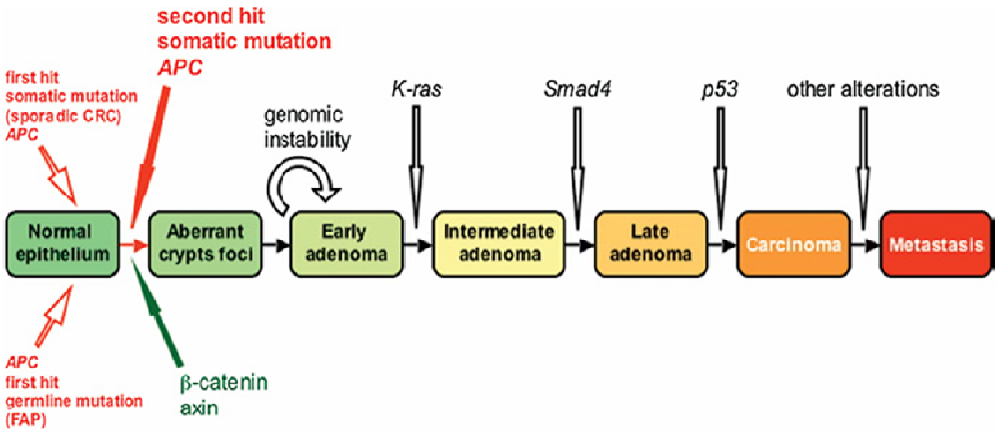


Figure 20. **Genetic model of colorectal carcinogenesis**
Mutations in *APC*, β-catenin or axin genes are required for tumour initiation. Subsequent progression towards malignancy is accompanied by genomic instability and sequential mutations in K-*ras*, *Smad4*, *p53* genes, as well as in other unknown genes. Adapted from reference #24.

Tumor progression is driven by the selection of specific genetic (and epigenetic) alterations that accumulate in strict order. In fact, whereas mutation events are random, the order in which they accumulate is non-random, supporting robustly the theory that only certain mutations confer selective advantage to a tumor to progress depending on its stage.

Adenomas are the precancerous lesions from which the malignant tumors arise. The malignant potential of an adenoma depends on its size, histologic type, and degree of atypia. Adenomatous polyps larger than 1 cm, with a large villous histologic component or with severe dysplasia have an increased frequency of malignant change. Only around 10-15% of these polyps progress to cancer and such progression generally takes around ten years. In MMR-deficient HNPCC patients, this progression period is reduced to approximately two years. Colorectal polyps are usually asymptomatic, frequently found during screening for colorectal cancer, and removed via colonoscopy. This procedure is adequate for most adenomas, including those with severe grade of dysplasia. The polyps which during exploration are found to be malignant may require surgical resection to ensure the complete elimination of cancer cells.

Colon cancer staging is an estimate of the condition of a particular cancer for diagnostic and research purposes. The systems for staging colorectal cancers largely depend on the extent of local

invasion, the degree of lymph node involvement, and whether there is distant metastasis. The most common currently used system for staging is the TNM system (Figure 21A). Since the number of combinations of categories is high, combinations are grouped to stages for better analysis (Figure 21B and C).

The aims for adopting a global standard staging system are to i) aid medical staff in staging the tumor and thus helping to plan the treatment, ii) give an indication of prognosis, iii) assist in the evaluation of the results of treatment, and iv) enable facilities around the world to collect information more productively.

A

T	Degree of invasion of the intestinal wall
T0	No evidence of tumor
Tis	Tumor *in situ* (present but without invasion
T1	Invasion through submucosa into lamina propria
T2	Invasion into the muscularis propria
T3	Invasion through the muscularis propria or to adjacent mucosa
T4	Invasion completely through the wall of the colon

N	Degree of lymphatic node involvement
N0	No lymph nodes involved
N1	One to three nodes involved
N2	Four or more nodes involved

M	Degree of metastasis
M0	No metastasis
M1	Metastasis present

B

Stage 0	Tis, N0, M0
Stage I	T1, N0, M0
	T2, N0, M0
Stage II	A: T3, N0, M0
	B: T4, N0, M0
Stage III	A: T1, N1, M0
	T2, N1, M0
	B: T3, N1, M0
	T4, N1, M0
	C: Any T, N2, M0
Stage IV	Any T, Any N, M1

C

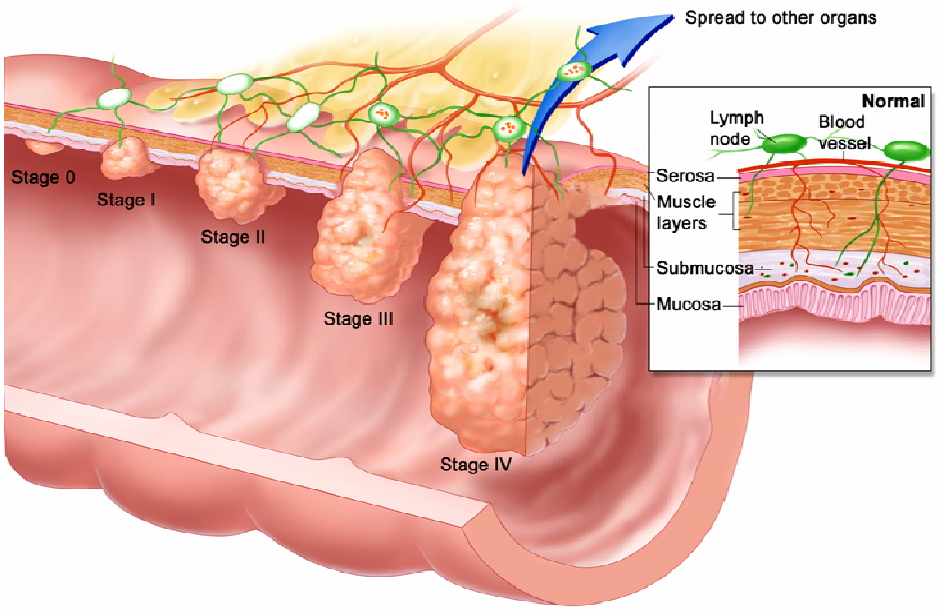


Figure 21. **Colorectal cancer staging system.** (A) The different categories of the TNM staging system, (B) Stages of the TNM system (based on grouped combinations of categories), (C) Graphic representation of different tumor staging through cancer progression.

References

1. Boyle P, Ferlay J. Cancer incidence and mortality in Europe, 2004. Ann Oncol 2005;16:481-8.
2. Jemal A, Murray T, Ward E, Samuels A, Tiwari RC, Ghafoor A, Feuer EJ, Thun MJ. Cancer statistics, 2005. CA Cancer J Clin 2005;55:10-30.
3. Hermiston ML, Wong MH, Gordon JI. Forced expression of E-cadherin in the mouse intestinal epithelium slows cell migration and provides evidence for nonautonomous regulation of cell fate in a self-renewing system. Genes Dev 1996;10:985-96.
4. Heath JP. Epithelial cell migration in the intestine. Cell Biol Int 1996;20:139-46.
5. Potten CS, Loeffler M. Stem cells: attributes, cycles, spirals, pitfalls and uncertainties. Lessons for and from the crypt. Development 1990;110:1001-20.
6. Marshman E, Booth C, Potten CS. The intestinal epithelial stem cell. Bioessays 2002;24:91-8.
7. Potten CS. Stem cells in gastrointestinal epithelium: numbers, characteristics and death. Philos Trans R Soc Lond B Biol Sci 1998;353:821-30.
8. Bach SP, Renehan AG, Potten CS. Stem cells: the intestinal stem cell as a paradigm. Carcinogenesis 2000;21:469-76.
9. Booth C, Potten CS. Gut instincts: thoughts on intestinal epithelial stem cells. J Clin Invest 2000;105:1493-9.
10. Bjerknes M, Cheng H. Clonal analysis of mouse intestinal epithelial progenitors. Gastroenterology 1999;116:7-14.
11. Nishimura S, Wakabayashi N, Toyoda K, Kashima K, Mitsufuji S. Expression of Musashi-1 in human normal colon crypt cells: a possible stem cell marker of human colon epithelium. Dig Dis Sci 2003;48:1523-9.
12. Potten CS, Booth C, Tudor GL, Booth D, Brady G, Hurley P, Ashton G, Clarke R, Sakakibara S, Okano H. Identification of a putative intestinal stem cell and early lineage marker; musashi-1. Differentiation 2003;71:28-41.
13. Stappenbeck TS, Mills JC, Gordon JI. Molecular features of adult mouse small intestinal epithelial progenitors. Proc Natl Acad Sci U S A 2003;100:1004-9.
14. Physilogy of the gastrointestinal tract. Raven Press Books, Ltd, 1987.
15. Text book of gastroenterology. Lippincott Williams & Wilkins, 1999.
16. Lewin K, Riddel, RH, Weinstein, WM. Gastrointestinal pathology and its clinical implications. IGAKU-SHOIN Ltd, 1992.
17. Delaunoit T, Neczyporenko F, Limburg PJ, Erlichman C. Pathogenesis and risk factors of small bowel adenocarcinoma: a colorectal cancer sibling? Am J Gastroenterol 2005;100:703-10.
18. Hanauer SB. Inflammatory bowel disease. N Engl J Med 1996;334:841-8.
19. Loftus EV, Jr., Schoenfeld P, Sandborn WJ. The epidemiology and natural history of Crohn's disease in population-based patient cohorts from North America: a systematic review. Aliment Pharmacol Ther 2002;16:51-60.
20. Bernstein CN, Wajda A, Svenson LW, MacKenzie A, Koehoorn M, Jackson M, Fedorak R, Israel D, Blanchard JF. The epidemiology of inflammatory bowel disease in Canada: a population-based study. Am J Gastroenterol 2006;101:1559-68.
21. Podolsky DK. Inflammatory bowel disease. N Engl J Med 2002;347:417-29.
22. O'Hara AM, Shanahan F. The gut flora as a forgotten organ. EMBO Rep 2006;7:688-93.
23. van Es JH, Clevers H. Notch and Wnt inhibitors as potential new drugs for intestinal neoplastic disease. Trends Mol Med 2005;11:496-502.
24. Pinto D, Clevers H. Wnt, stem cells and cancer in the intestine. Biol Cell 2005;97:185-96.
25. Ilyas M, Straub J, Tomlinson IP, Bodmer WF. Genetic pathways in colorectal and other cancers. Eur J Cancer 1999;35:335-51.
26. Sancho E, Batlle E, Clevers H. Signaling pathways in intestinal development and cancer. Annu Rev Cell Dev Biol 2004;20:695-723.
27. Polakis P. Wnt signaling and cancer. Genes Dev 2000;14:1837-51.
28. Moon RT, Kohn AD, De Ferrari GV, Kaykas A. WNT and beta-catenin signalling: diseases and therapies. Nat Rev Genet 2004;5:691-701.
29. Artavanis-Tsakonas S, Rand MD, Lake RJ. Notch signaling: cell fate control and signal integration in development. Science 1999;284:770-6.
30. Lefort K, Dotto GP. Notch signaling in the integrated control of keratinocyte growth/differentiation and tumor suppression. Semin Cancer Biol 2004;14:374-86.

31. Massague J, Blain SW, Lo RS. TGFbeta signaling in growth control, cancer, and heritable disorders. Cell 2000;103:295-309.
32. Blobe GC, Schiemann WP, Lodish HF. Role of transforming growth factor beta in human disease. N Engl J Med 2000;342:1350-8.
33. Waite KA, Eng C. From developmental disorder to heritable cancer: it's all in the BMP/TGF-beta family. Nat Rev Genet 2003;4:763-73.
34. Zhang J, Li L. BMP signaling and stem cell regulation. Dev Biol 2005;284:1-11.
35. Bitgood MJ, McMahon AP. Hedgehog and Bmp genes are coexpressed at many diverse sites of cell-cell interaction in the mouse embryo. Dev Biol 1995;172:126-38.
36. Bijlsma MF, Spek CA, Peppelenbosch MP. Hedgehog: an unusual signal transducer. Bioessays 2004;26:387-94.
37. Hancock JF. Ras proteins: different signals from different locations. Nat Rev Mol Cell Biol 2003;4:373-84.
38. Malumbres M, Barbacid M. RAS oncogenes: the first 30 years. Nat Rev Cancer 2003;3:459-65.
39. Pires-daSilva A, Sommer RJ. The evolution of signalling pathways in animal development. Nat Rev Genet 2003;4:39-49.
40. Korinek V, Barker N, Morin PJ, van Wichen D, de Weger R, Kinzler KW, Vogelstein B, Clevers H. Constitutive transcriptional activation by a beta-catenin-Tcf complex in APC-/- colon carcinoma. Science 1997;275:1784-7.
41. Morin PJ, Sparks AB, Korinek V, Barker N, Clevers H, Vogelstein B, Kinzler KW. Activation of beta-catenin-Tcf signaling in colon cancer by mutations in beta-catenin or APC. Science 1997;275:1787-90.
42. Bienz M, Clevers H. Linking colorectal cancer to Wnt signaling. Cell 2000;103:311-20.
43. Booth C, Brady G, Potten CS. Crowd control in the crypt. Nat Med 2002;8:1360-1.
44. Kinzler KW, Vogelstein B. Lessons from hereditary colorectal cancer. Cell 1996;87:159-70.
45. Korinek V, Barker N, Moerer P, van Donselaar E, Huls G, Peters PJ, Clevers H. Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4. Nat Genet 1998;19:379-83.
46. Batlle E, Henderson JT, Beghtel H, van den Born MM, Sancho E, Huls G, Meeldijk J, Robertson J, van de Wetering M, Pawson T, Clevers H. Beta-catenin and TCF mediate cell positioning in the intestinal epithelium by controlling the expression of EphB/ephrinB. Cell 2002;111:251-63.
47. van de Wetering M, Sancho E, Verweij C, de Lau W, Oving I, Hurlstone A, van der Horn K, Batlle E, Coudreuse D, Haramis AP, Tjon-Pon-Fong M, Moerer P, van den Born M, Soete G, Pals S, Eilers M, Medema R, Clevers H. The beta-catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. Cell 2002;111:241-50.
48. Pinto D, Gregorieff A, Begthel H, Clevers H. Canonical Wnt signals are essential for homeostasis of the intestinal epithelium. Genes Dev 2003;17:1709-13.
49. Munemitsu S, Albert I, Souza B, Rubinfeld B, Polakis P. Regulation of intracellular beta-catenin levels by the adenomatous polyposis coli (APC) tumor-suppressor protein. Proc Natl Acad Sci U S A 1995;92:3046-50.
50. Rubinfeld B, Albert I, Porfiri E, Fiol C, Munemitsu S, Polakis P. Binding of GSK3beta to the APC-beta-catenin complex and regulation of complex assembly. Science 1996;272:1023-6.
51. Rubinfeld B, Robbins P, El-Gamil M, Albert I, Porfiri E, Polakis P. Stabilization of beta-catenin by genetic defects in melanoma cell lines. Science 1997;275:1790-2.
52. Van der Flier L, Sabates-Bellver, J, Oving, I, Haegebarth, A, de Palo, M, Anti, M, van Gijn, ME, Suijkerbuijk, S, Van de Wetering, M, Marra, G, Clevers, H. The intestinal Wnt/TCF signature. Gastroenterology DOI: 10.1053/j.gastro.2006.08.039 2006.
53. Moser AR, Pitot HC, Dove WF. A dominant mutation that predisposes to multiple intestinal neoplasia in the mouse. Science 1990;247:322-4.
54. Oshima M, Oshima H, Kitagawa K, Kobayashi M, Itakura C, Taketo M. Loss of Apc heterozygosity and abnormal tissue building in nascent intestinal polyps in mice carrying a truncated Apc gene. Proc Natl Acad Sci U S A 1995;92:4482-6.
55. Su LK, Kinzler KW, Vogelstein B, Preisinger AC, Moser AR, Luongo C, Gould KA, Dove WF. Multiple intestinal neoplasia caused by a mutation in the murine homolog of the APC gene. Science 1992;256:668-70.
56. Mumm JS, Schroeter EH, Saxena MT, Griesemer A, Tian X, Pan DJ, Ray WJ, Kopan R. A ligand-induced extracellular cleavage regulates gamma-secretase-like proteolytic activation of Notch1. Mol Cell 2000;5:197-206.

57. Baron M. An overview of the Notch signalling pathway. Semin Cell Dev Biol 2003;14:113-9.
58. Heitzler P, Bourouis M, Ruel L, Carteret C, Simpson P. Genes of the Enhancer of split and achaete-scute complexes are required for a regulatory loop between Notch and Delta during lateral signalling in Drosophila. Development 1996;122:161-71.
59. Oellers N, Dehio M, Knust E. bHLH proteins encoded by the Enhancer of split complex of Drosophila negatively interfere with transcriptional activation mediated by proneural genes. Mol Gen Genet 1994;244:465-73.
60. Schroder N, Gossler A. Expression of Notch pathway components in fetal and adult mouse small intestine. Gene Expr Patterns 2002;2:247-50.
61. Jensen J, Pedersen EE, Galante P, Hald J, Heller RS, Ishibashi M, Kageyama R, Guillemot F, Serup P, Madsen OD. Control of endodermal endocrine development by Hes-1. Nat Genet 2000;24:36-44.
62. Apelqvist A, Li H, Sommer L, Beatus P, Anderson DJ, Honjo T, Hrabe de Angelis M, Lendahl U, Edlund H. Notch signalling controls pancreatic cell differentiation. Nature 1999;400:877-81.
63. De Strooper B, Annaert W, Cupers P, Saftig P, Craessaerts K, Mumm JS, Schroeter EH, Schrijvers V, Wolfe MS, Ray WJ, Goate A, Kopan R. A presenilin-1-dependent gamma-secretase-like protease mediates release of Notch intracellular domain. Nature 1999;398:518-22.
64. Jenny M, Uhl C, Roche C, Duluc I, Guillermin V, Guillemot F, Jensen J, Kedinger M, Gradwohl G. Neurogenin3 is differentially required for endocrine cell fate specification in the intestinal and gastric epithelium. Embo J 2002;21:6338-47.
65. Kopan R, Goate A. A common enzyme connects notch signaling and Alzheimer's disease. Genes Dev 2000;14:2799-806.
66. Naya FJ, Huang HP, Qiu Y, Mutoh H, DeMayo FJ, Leiter AB, Tsai MJ. Diabetes, defective pancreatic morphogenesis, and abnormal enteroendocrine differentiation in BETA2/neuroD-deficient mice. Genes Dev 1997;11:2323-34.
67. Wong GT, Manfra D, Poulet FM, Zhang Q, Josien H, Bara T, Engstrom L, Pinzon-Ortiz M, Fine JS, Lee HJ, Zhang L, Higgins GA, Parker EM. Chronic treatment with the gamma-secretase inhibitor LY-411,575 inhibits beta-amyloid peptide production and alters lymphopoiesis and intestinal cell differentiation. J Biol Chem 2004;279:12876-82.
68. Yang Q, Bermingham NA, Finegold MJ, Zoghbi HY. Requirement of Math1 for secretory cell lineage commitment in the mouse intestine. Science 2001;294:2155-8.
69. Shi Y, Massague J. Mechanisms of TGF-beta signaling from cell membrane to the nucleus. Cell 2003;113:685-700.
70. Massague J. How cells read TGF-beta signals. Nat Rev Mol Cell Biol 2000;1:169-78.
71. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. Cell 1990;61:759-67.
72. Grady WM, Myeroff LL, Swinler SE, Rajput A, Thiagalingam S, Lutterbaugh JD, Neumann A, Brattain MG, Chang J, Kim SJ, Kinzler KW, Vogelstein B, Willson JK, Markowitz S. Mutational inactivation of transforming growth factor beta receptor type II in microsatellite stable colon cancers. Cancer Res 1999;59:320-4.
73. Eppert K, Scherer SW, Ozcelik H, Pirone R, Hoodless P, Kim H, Tsui LC, Bapat B, Gallinger S, Andrulis IL, Thomsen GH, Wrana JL, Attisano L. MADR2 maps to 18q21 and encodes a TGFbeta-regulated MAD-related protein that is functionally mutated in colorectal carcinoma. Cell 1996;86:543-52.
74. Takagi Y, Kohmura H, Futamura M, Kida H, Tanemura H, Shimokawa K, Saji S. Somatic alterations of the DPC4 gene in human colorectal cancers in vivo. Gastroenterology 1996;111:1369-72.
75. Weinstein M, Yang X, Deng C. Functions of mammalian Smad genes as revealed by targeted gene disruption in mice. Cytokine Growth Factor Rev 2000;11:49-58.
76. Dunker N, Krieglstein K. Targeted mutations of transforming growth factor-beta genes reveal important roles in mouse development and adult homeostasis. Eur J Biochem 2000;267:6982-8.
77. Takaku K, Oshima M, Miyoshi H, Matsui M, Seldin MF, Taketo MM. Intestinal tumorigenesis in compound mutant mice of both Dpc4 (Smad4) and Apc genes. Cell 1998;92:645-56.
78. Zwijsen A, Verschueren K, Huylebroeck D. New intracellular components of bone morphogenetic protein/Smad signaling cascades. FEBS Lett 2003;546:133-9.
79. Derynck R, Zhang YE. Smad-dependent and Smad-independent pathways in TGF-beta family signalling. Nature 2003;425:577-84.
80. Massague J. Integration of Smad and MAPK pathways: a link and a linker revisited. Genes Dev 2003;17:2993-7.

81. Yamaguchi K, Nagai S, Ninomiya-Tsuji J, Nishita M, Tamai K, Irie K, Ueno N, Nishida E, Shibuya H, Matsumoto K. XIAP, a cellular member of the inhibitor of apoptosis protein family, links the receptors to TAB1-TAK1 in the BMP signaling pathway. Embo J 1999;18:179-87.
82. Takaesu G, Surabhi RM, Park KJ, Ninomiya-Tsuji J, Matsumoto K, Gaynor RB. TAK1 is critical for IkappaB kinase-mediated activation of the NF-kappaB pathway. J Mol Biol 2003;326:105-15.
83. Aubin J, Davy A, Soriano P. In vivo convergence of BMP and MAPK signaling pathways: impact of differential Smad1 phosphorylation on development and homeostasis. Genes Dev 2004;18:1482-94.
84. Howe JR, Bair JL, Sayed MG, Anderson ME, Mitros FA, Petersen GM, Velculescu VE, Traverso G, Vogelstein B. Germline mutations of the gene encoding bone morphogenetic protein receptor 1A in juvenile polyposis. Nat Genet 2001;28:184-7.
85. Sayed MG, Ahmed AF, Ringold JR, Anderson ME, Bair JL, Mitros FA, Lynch HT, Tinley ST, Petersen GM, Giardiello FM, Vogelstein B, Howe JR. Germline SMAD4 or BMPR1A mutations and phenotype of juvenile polyposis. Ann Surg Oncol 2002;9:901-6.
86. Zhou XP, Woodford-Richens K, Lehtonen R, Kurose K, Aldred M, Hampel H, Launonen V, Virta S, Pilarski R, Salovaara R, Bodmer WF, Conrad BA, Dunlop M, Hodgson SV, Iwama T, Jarvinen H, Kellokumpu I, Kim JC, Leggett B, Markie D, Mecklin JP, Neale K, Phillips R, Piris J, Rozen P, Houlston RS, Aaltonen LA, Tomlinson IP, Eng C. Germline mutations in BMPR1A/ALK3 cause a subset of cases of juvenile polyposis syndrome and of Cowden and Bannayan-Riley-Ruvalcaba syndromes. Am J Hum Genet 2001;69:704-11.
87. Haramis AP, Begthel H, van den Born M, van Es J, Jonkheer S, Offerhaus GJ, Clevers H. De novo crypt formation and juvenile polyposis on BMP inhibition in mouse intestine. Science 2004;303:1684-6.
88. Collins RT, Cohen SM. A genetic screen in Drosophila for identifying novel components of the hedgehog signaling pathway. Genetics 2005;170:173-84.
89. Mohler J. Requirements for hedgehod, a segmental polarity gene, in patterning larval and adult cuticle of Drosophila. Genetics 1988;120:1061-72.
90. Lum L, Beachy PA. The Hedgehog response network: sensors, switches, and routers. Science 2004;304:1755-9.
91. Chen W, Ren XR, Nelson CD, Barak LS, Chen JK, Beachy PA, de Sauvage F, Lefkowitz RJ. Activity-dependent internalization of smoothened mediated by beta-arrestin 2 and GRK2. Science 2004;306:2257-60.
92. Ingham PW, McMahon AP. Hedgehog signaling in animal development: paradigms and principles. Genes Dev 2001;15:3059-87.
93. Nybakken K, Perrimon N. Hedgehog signal transduction: recent findings. Curr Opin Genet Dev 2002;12:503-11.
94. Alcedo J, Zou Y, Noll M. Posttranscriptional regulation of smoothened is part of a self-correcting mechanism in the Hedgehog signaling system. Mol Cell 2000;6:457-65.
95. Mo R, Kim JH, Zhang J, Chiang C, Hui CC, Kim PC. Anorectal malformations caused by defects in sonic hedgehog signaling. Am J Pathol 2001;159:765-74.
96. Kimmel SG, Mo R, Hui CC, Kim PC. New mouse models of congenital anorectal malformations. J Pediatr Surg 2000;35:227-30; discussion 230-1.
97. Ramalho-Santos M, Melton DA, McMahon AP. Hedgehog signals regulate multiple aspects of gastrointestinal development. Development 2000;127:2763-72.
98. Pasca di Magliano M, Hebrok M. Hedgehog signalling in cancer formation and maintenance. Nat Rev Cancer 2003;3:903-11.
99. Wetmore C. Sonic hedgehog in normal and neoplastic proliferation: insight gained from human tumors and animal models. Curr Opin Genet Dev 2003;13:34-42.
100. Berman DM, Karhadkar SS, Maitra A, Montes De Oca R, Gerstenblith MR, Briggs K, Parker AR, Shimada Y, Eshleman JR, Watkins DN, Beachy PA. Widespread requirement for Hedgehog ligand stimulation in growth of digestive tract tumours. Nature 2003;425:846-51.
101. Thayer SP, di Magliano MP, Heiser PW, Nielsen CM, Roberts DJ, Lauwers GY, Qi YP, Gysin S, Fernandez-del Castillo C, Yajnik V, Antoniu B, McMahon M, Warshaw AL, Hebrok M. Hedgehog is an early and late mediator of pancreatic cancer tumorigenesis. Nature 2003;425:851-6.
102. Vogelstein B, Fearon ER, Hamilton SR, Kern SE, Preisinger AC, Leppert M, Nakamura Y, White R, Smits AM, Bos JL. Genetic alterations during colorectal-tumor development. N Engl J Med 1988;319:525-32.

103. Andreyev HJ, Norman AR, Cunningham D, Oates J, Dix BR, Iacopetta BJ, Young J, Walsh T, Ward R, Hawkins N, Beranek M, Jandik P, Benamouzig R, Jullian E, Laurent-Puig P, Olschwang S, Muller O, Hoffmann I, Rabes HM, Zietz C, Troungos C, Valavanis C, Yuen ST, Ho JW, Croke CT, O'Donoghue DP, Giaretti W, Rapallo A, Russo A, Bazan V, Tanaka M, Omura K, Azuma T, Ohkusa T, Fujimori T, Ono Y, Pauly M, Faber C, Glaesener R, de Goeij AF, Arends JW, Andersen SN, Lovig T, Breivik J, Gaudernack G, Clausen OP, De Angelis PD, Meling GI, Rognum TO, Smith R, Goh HS, Font A, Rosell R, Sun XF, Zhang H, Benhattar J, Losi L, Lee JQ, Wang ST, Clarke PA, Bell S, Quirke P, Bubb VJ, Piris J, Cruickshank NR, Morton D, Fox JC, Al-Mulla F, Lees N, Hall CN, Snary D, Wilkinson K, Dillon D, Costa J, Pricolo VE, Finkelstein SD, Thebo JS, Senagore AJ, Halter SA, Wadler S, Malik S, Krtolica K, Urosevic N. Kirsten ras mutations in patients with colorectal cancer: the 'RASCAL II' study. Br J Cancer 2001;85:692-6.
104. Davies H, Bignell GR, Cox C, Stephens P, Edkins S, Clegg S, Teague J, Woffendin H, Garnett MJ, Bottomley W, Davis N, Dicks E, Ewing R, Floyd Y, Gray K, Hall S, Hawes R, Hughes J, Kosmidou V, Menzies A, Mould C, Parker A, Stevens C, Watt S, Hooper S, Wilson R, Jayatilake H, Gusterson BA, Cooper C, Shipley J, Hargrave D, Pritchard-Jones K, Maitland N, Chenevix-Trench G, Riggins GJ, Bigner DD, Palmieri G, Cossu A, Flanagan A, Nicholson A, Ho JW, Leung SY, Yuen ST, Weber BL, Seigler HF, Darrow TL, Paterson H, Marais R, Marshall CJ, Wooster R, Stratton MR, Futreal PA. Mutations of the BRAF gene in human cancer. Nature 2002;417:949-54.
105. Rajagopalan H, Bardelli A, Lengauer C, Kinzler KW, Vogelstein B, Velculescu VE. Tumorigenesis: RAF/RAS oncogenes and mismatch-repair status. Nature 2002;418:934.
106. Coopersmith CM, Chandrasekaran C, McNevin MS, Gordon JI. Bi-transgenic mice reveal that K-rasVal12 augments a p53-independent apoptosis when small intestinal villus enterocytes reenter the cell cycle. J Cell Biol 1997;138:167-79.
107. Kim SH, Roth KA, Moser AR, Gordon JI. Transgenic mouse models that explore the multistep hypothesis of intestinal neoplasia. J Cell Biol 1993;123:877-93.
108. Janssen KP, el-Marjou F, Pinto D, Sastre X, Rouillard D, Fouquet C, Soussi T, Louvard D, Robine S. Targeted expression of oncogenic K-ras in intestinal epithelium causes spontaneous tumorigenesis in mice. Gastroenterology 2002;123:492-504.
109. Johnson L, Mercer K, Greenbaum D, Bronson RT, Crowley D, Tuveson DA, Jacks T. Somatic activation of the K-ras oncogene causes early onset lung cancer in mice. Nature 2001;410:1111-6.
110. Tamaoki N. The rasH2 transgenic mouse: nature of the model and mechanistic studies on tumorigenesis. Toxicol Pathol 2001;29 Suppl:81-9.
111. Lodish H BA, Matsudaira P, Kaiser CA, Krieger M, Scott MP, Zipursky SL, Darnell J. . Molecular Biology of the Cell., 2004.
112. Lindahl T, Wood RD. Quality control by DNA repair. Science 1999;286:1897-905.
113. Lindahl T. Suppression of spontaneous mutagenesis in human cells by DNA base excision-repair. Mutat Res 2000;462:129-35.
114. Peterson CL, Cote J. Cellular machineries for chromosomal DNA repair. Genes Dev 2004;18:602-16.
115. Watson JD BT, Bell SP, Gann A, Levine M, Losick R. Molecular Biology of the Gene, 2004.
116. Kunkel TA, Bebenek K. DNA replication fidelity. Annu Rev Biochem 2000;69:497-529.
117. Jiricny J. The multifaceted mismatch-repair system. Nat Rev Mol Cell Biol 2006;7:335-46.
118. Kunkel TA, Erie DA. DNA mismatch repair. Annu Rev Biochem 2005;74:681-710.
119. Modrich P. Mechanisms in eukaryotic mismatch repair. J Biol Chem 2006;281:30305-9.
120. Lahue RS, Au KG, Modrich P. DNA mismatch correction in a defined system. Science 1989;245:160-4.
121. Marra G, Jiricny J. Genome Instability in Cancer Development. 85-123, 2005.
122. Cannavo E, Marra G, Sabates-Bellver J, Menigatti M, Lipkin SM, Fischer F, Cejka P, Jiricny J. Expression of the MutL homologue hMLH3 in human cells and its role in DNA mismatch repair. Cancer Res 2005;65:10759-66.
123. Chen PC, Dudley S, Hagen W, Dizon D, Paxton L, Reichow D, Yoon SR, Yang K, Arnheim N, Liskay RM, Lipkin SM. Contributions by MutL homologues Mlh3 and Pms2 to DNA mismatch repair and tumor suppression in the mouse. Cancer Res 2005;65:8662-70.
124. Lipkin SM, Moens PB, Wang V, Lenzi M, Shanmugarajah D, Gilgeous A, Thomas J, Cheng J, Touchman JW, Green ED, Schwartzberg P, Collins FS, Cohen PE. Meiotic arrest and aneuploidy in MLH3-deficient mice. Nat Genet 2002;31:385-90.

125. Bruni R, Martin D, Jiricny J. d(GATC) sequences influence Escherichia coli mismatch repair in a distance-dependent manner from positions both upstream and downstream of the mismatch. Nucleic Acids Res 1988;16:4875-90.
126. Kramer B, Kramer W, Fritz HJ. Different base/base mismatches are corrected with different efficiencies by the methyl-directed DNA mismatch-repair system of E. coli. Cell 1984;38:879-87.
127. Fang WH, Modrich P. Human strand-specific mismatch repair occurs by a bidirectional mechanism similar to that of the bacterial reaction. J Biol Chem 1993;268:11838-44.
128. Holmes J, Jr., Clark S, Modrich P. Strand-specific mismatch correction in nuclear extracts of human and Drosophila melanogaster cell lines. Proc Natl Acad Sci U S A 1990;87:5837-41.
129. Thomas DC, Roberts JD, Kunkel TA. Heteroduplex repair in extracts of human HeLa cells. J Biol Chem 1991;266:3744-51.
130. Gradia S, Subramanian D, Wilson T, Acharya S, Makhov A, Griffith J, Fishel R. hMSH2-hMSH6 forms a hydrolysis-independent sliding clamp on mismatched DNA. Mol Cell 1999;3:255-61.
131. Blackwell LJ, Martik D, Bjornson KP, Bjornson ES, Modrich P. Nucleotide-promoted release of hMutSalpha from heteroduplex DNA is consistent with an ATP-dependent translocation mechanism. J Biol Chem 1998;273:32055-62.
132. Martik D, Baitinger C, Modrich P. Differential specificities and simultaneous occupancy of human MutSalpha nucleotide binding sites. J Biol Chem 2004;279:28402-10.
133. Wang H, Hays JB. Mismatch repair in human nuclear extracts: effects of internal DNA-hairpin structures between mismatches and excision-initiation nicks on mismatch correction and mismatch-provoked excision. J Biol Chem 2003;278:28686-93.
134. Wang H, Hays JB. Signaling from DNA mispairs to mismatch-repair excision sites despite intervening blockades. Embo J 2004;23:2126-33.
135. Blake C, Tsao JL, Wu A, Shibata D. Stepwise deletions of polyA sequences in mismatch repair-deficient colorectal cancers. Am J Pathol 2001;158:1867-70.
136. Lengauer C, Kinzler KW, Vogelstein B. Genetic instabilities in human cancers. Nature 1998;396:643-9.
137. Boland CR, Thibodeau SN, Hamilton SR, Sidransky D, Eshleman JR, Burt RW, Meltzer SJ, Rodriguez-Bigas MA, Fodde R, Ranzani GN, Srivastava S. A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. Cancer Res 1998;58:5248-57.
138. Eshleman JR, Lang EZ, Bowerfind GK, Parsons R, Vogelstein B, Willson JK, Veigl ML, Sedwick WD, Markowitz SD. Increased mutation rate at the hprt locus accompanies microsatellite instability in colon cancer. Oncogene 1995;10:33-7.
139. Ionov Y, Peinado MA, Malkhosyan S, Shibata D, Perucho M. Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. Nature 1993;363:558-61.
140. Ciotta C, Ceccotti S, Aquilina G, Humbert O, Palombo F, Jiricny J, Bignami M. Increased somatic recombination in methylation tolerant human cells with defective DNA mismatch repair. J Mol Biol 1998;276:705-19.
141. Abdel-Rahman WM, Katsura K, Rens W, Gorman PA, Sheer D, Bicknell D, Bodmer WF, Arends MJ, Wyllie AH, Edwards PA. Spectral karyotyping suggests additional subsets of colorectal cancers characterized by pattern of chromosome rearrangement. Proc Natl Acad Sci U S A 2001;98:2538-43.
142. Markowitz S, Wang J, Myeroff L, Parsons R, Sun L, Lutterbaugh J, Fan RS, Zborowska E, Kinzler KW, Vogelstein B, et al. Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability. Science 1995;268:1336-8.
143. Rampino N, Yamamoto H, Ionov Y, Li Y, Sawai H, Reed JC, Perucho M. Somatic frameshift mutations in the BAX gene in colon cancers of the microsatellite mutator phenotype. Science 1997;275:967-9.
144. Duval A, Iacopetta B, Ranzani GN, Lothe RA, Thomas G, Hamelin R. Variable mutation frequencies in coding repeats of TCF-4 and other target genes in colon, gastric and endometrial carcinoma showing microsatellite instability. Oncogene 1999;18:6806-9.
145. Liu W, Dong X, Mai M, Seelan RS, Taniguchi K, Krishnadath KK, Halling KC, Cunningham JM, Boardman LA, Qian C, Christensen E, Schmidt SS, Roche PC, Smith DI, Thibodeau SN. Mutations in AXIN2 cause colorectal cancer with defective mismatch repair by activating beta-catenin/TCF signalling. Nat Genet 2000;26:146-7.

146. Peltomaki P. Role of DNA mismatch repair defects in the pathogenesis of human cancer. J Clin Oncol 2003;21:1174-9.
147. Lynch HT, de la Chapelle A. Hereditary colorectal cancer. N Engl J Med 2003;348:919-32.
148. de la Chapelle A, Peltomaki P. Genetics of hereditary colon cancer. Annu Rev Genet 1995;29:329-48.
149. de la Chapelle A, Peltomaki P. The genetics of hereditary common cancers. Curr Opin Genet Dev 1998;8:298-303.
150. Hampel H, Peltomaki P. Hereditary colorectal cancer: risk assessment and management. Clin Genet 2000;58:89-97.
151. Lynch HT, de la Chapelle A. Genetic susceptibility to non-polyposis colorectal cancer. J Med Genet 1999;36:801-18.
152. Peltomaki P. Microsatellite instability and hereditary non-polyposis colon cancer. J Pathol 1995;176:329-30.
153. Peltomaki P. Microsatellite instability as an indicator of hereditary susceptibility to colon cancer. Gastroenterology 1995;109:2031-3.
154. Peltomaki P. Deficient DNA mismatch repair: a common etiologic factor for colon cancer. Hum Mol Genet 2001;10:735-40.
155. Peltomaki P. DNA mismatch repair and cancer. Mutat Res 2001;488:77-85.
156. Peltomaki P. Lynch syndrome genes. Fam Cancer 2005;4:227-32.
157. Radtke F, Clevers H. Self-renewal and cancer of the gut: two sides of a coin. Science 2005;307:1904-9.
158. Vasen HF. Colorectal cancer and family history. Ann Chir Gynaecol 2000;89:179-84.
159. Vasen HF. Clinical description of the Lynch syndrome [hereditary nonpolyposis colorectal cancer (HNPCC)]. Fam Cancer 2005;4:219-25.
160. Vasen HF, Mecklin JP, Khan PM, Lynch HT. The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC). Dis Colon Rectum 1991;34:424-5.
161. Vasen HF, Watson P, Mecklin JP, Lynch HT. New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC. Gastroenterology 1999;116:1453-6.
162. Knudson AG. Antioncogenes and human cancer. Proc Natl Acad Sci U S A 1993;90:10914-21.
163. Rodriguez-Bigas MA, Boland CR, Hamilton SR, Henson DE, Jass JR, Khan PM, Lynch H, Perucho M, Smyrk T, Sobin L, Srivastava S. A National Cancer Institute Workshop on Hereditary Nonpolyposis Colorectal Cancer Syndrome: meeting highlights and Bethesda guidelines. J Natl Cancer Inst 1997;89:1758-62.
164. Umar A, Boland CR, Terdiman JP, Syngal S, de la Chapelle A, Ruschoff J, Fishel R, Lindor NM, Burgart LJ, Hamelin R, Hamilton SR, Hiatt RA, Jass J, Lindblom A, Lynch HT, Peltomaki P, Ramsey SD, Rodriguez-Bigas MA, Vasen HF, Hawk ET, Barrett JC, Freedman AN, Srivastava S. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. J Natl Cancer Inst 2004;96:261-8.
165. Syngal S, Fox EA, Eng C, Kolodner RD, Garber JE. Sensitivity and specificity of clinical criteria for hereditary non-polyposis colorectal cancer associated mutations in MSH2 and MLH1. J Med Genet 2000;37:641-5.
166. Wullenweber HP, Sutter C, Autschbach F, Willeke F, Kienle P, Benner A, Bahring J, Kadmon M, Herfarth C, von Knebel Doeberitz M, Gebert J. Evaluation of Bethesda guidelines in relation to microsatellite instability. Dis Colon Rectum 2001;44:1281-9.
167. Jass JR, Stewart SM. Evolution of hereditary non-polyposis colorectal cancer. Gut 1992;33:783-6.
168. Nystrom-Lahti M, Wu Y, Moisio AL, Hofstra RM, Osinga J, Mecklin JP, Jarvinen HJ, Leisti J, Buys CH, de la Chapelle A, Peltomaki P. DNA mismatch repair gene mutations in 55 kindreds with verified or putative hereditary non-polyposis colorectal cancer. Hum Mol Genet 1996;5:763-9.
169. Wijnen J, de Leeuw W, Vasen H, van der Klift H, Moller P, Stormorken A, Meijers-Heijboer H, Lindhout D, Menko F, Vossen S, Moslein G, Tops C, Brocker-Vriends A, Wu Y, Hofstra R, Sijmons R, Cornelisse C, Morreau H, Fodde R. Familial endometrial cancer in female carriers of MSH6 germline mutations. Nat Genet 1999;23:142-4.
170. Wijnen J, Khan PM, Vasen H, van der Klift H, Mulder A, van Leeuwen-Cornelisse I, Bakker B, Losekoot M, Moller P, Fodde R. Hereditary nonpolyposis colorectal cancer families not

complying with the Amsterdam criteria show extremely low frequency of mismatch-repair-gene mutations. Am J Hum Genet 1997;61:329-35.
171. Truninger K, Menigatti M, Luz J, Russell A, Haider R, Gebbers JO, Bannwart F, Yurtsever H, Neuweiler J, Riehle HM, Cattaruzza MS, Heinimann K, Schar P, Jiricny J, Marra G. Immunohistochemical analysis reveals high frequency of PMS2 defects in colorectal cancer. Gastroenterology 2005;128:1160-71.
172. Wu Y, Berends MJ, Sijmons RH, Mensink RG, Verlind E, Kooi KA, van der Sluis T, Kempinga C, van dDer Zee AG, Hollema H, Buys CH, Kleibeuker JH, Hofstra RM. A role for MLH3 in hereditary nonpolyposis colorectal cancer. Nat Genet 2001;29:137-8.
173. Souza RF, Wang S, Thakar M, Smolinski KN, Yin J, Zou TT, Kong D, Abraham JM, Toretsky JA, Meltzer SJ. Expression of the wild-type insulin-like growth factor II receptor gene suppresses growth and causes death in colorectal carcinoma cells. Oncogene 1999;18:4063-8.
174. Duval A, Gayet J, Zhou XP, Iacopetta B, Thomas G, Hamelin R. Frequent frameshift mutations of the TCF-4 gene in colorectal cancers with microsatellite instability. Cancer Res 1999;59:4213-5.
175. Laghi L, Ranzani GN, Bianchi P, Mori A, Heinimann K, Orbetegli O, Spaudo MR, Luinetti O, Francisconi S, Roncalli M, Solcia E, Malesci A. Frameshift mutations of human gastrin receptor gene (hGARE) in gastrointestinal cancers with microsatellite instability. Lab Invest 2002;82:265-71.
176. Bicknell DC, Kaklamanis L, Hampson R, Bodmer WF, Karran P. Selection for beta 2-microglobulin mutation in mismatch repair-defective colorectal carcinomas. Curr Biol 1996;6:1695-7.
177. Chung DC, Rustgi AK. The hereditary nonpolyposis colorectal cancer syndrome: genetics and clinical implications. Ann Intern Med 2003;138:560-70.
178. Lawes DA, SenGupta S, Boulos PB. The clinical importance and prognostic implications of microsatellite instability in sporadic cancer. Eur J Surg Oncol 2003;29:201-12.
179. Kuismanen SA, Holmberg MT, Salovaara R, Schweizer P, Aaltonen LA, de La Chapelle A, Nystrom-Lahti M, Peltomaki P. Epigenetic phenotypes distinguish microsatellite-stable and -unstable colorectal cancers. Proc Natl Acad Sci U S A 1999;96:12661-6.
180. Veigl ML, Kasturi L, Olechnowicz J, Ma AH, Lutterbaugh JD, Periyasamy S, Li GM, Drummond J, Modrich PL, Sedwick WD, Markowitz SD. Biallelic inactivation of hMLH1 by epigenetic gene silencing, a novel mechanism causing human MSI cancers. Proc Natl Acad Sci U S A 1998;95:8698-702.
181. Ahuja N, Li Q, Mohan AL, Baylin SB, Issa JP. Aging and DNA methylation in colorectal mucosa and cancer. Cancer Res 1998;58:5489-94.
182. Toyota M, Ahuja N, Ohe-Toyota M, Herman JG, Baylin SB, Issa JP. CpG island methylator phenotype in colorectal cancer. Proc Natl Acad Sci U S A 1999;96:8681-6.
183. Gryfe R, Kim H, Hsieh ET, Aronson MD, Holowaty EJ, Bull SB, Redston M, Gallinger S. Tumor microsatellite instability and clinical outcome in young patients with colorectal cancer. N Engl J Med 2000;342:69-77.
184. Kim H, Jen J, Vogelstein B, Hamilton SR. Clinical and pathological characteristics of sporadic colorectal carcinomas with DNA replication errors in microsatellite sequences. Am J Pathol 1994;145:148-56.
185. Lothe RA, Peltomaki P, Meling GI, Aaltonen LA, Nystrom-Lahti M, Pylkkanen L, Heimdal K, Andersen TI, Moller P, Rognum TO, et al. Genomic instability in colorectal cancer: relationship to clinicopathological variables and family history. Cancer Res 1993;53:5849-52.
186. Malkhosyan SR, Yamamoto H, Piao Z, Perucho M. Late onset and high incidence of colon cancer of the mutator phenotype with hypermethylated hMLH1 gene in women. Gastroenterology 2000;119:598.
187. Cunningham JM, Kim CY, Christensen ER, Tester DJ, Parc Y, Burgart LJ, Halling KC, McDonnell SK, Schaid DJ, Walsh Vockley C, Kubly V, Nelson H, Michels VV, Thibodeau SN. The frequency of hereditary defective mismatch repair in a prospective series of unselected colorectal carcinomas. Am J Hum Genet 2001;69:780-90.
188. Percesepe A, Kristo P, Aaltonen LA, Ponz de Leon M, de la Chapelle A, Peltomaki P. Mismatch repair genes and mononucleotide tracts as mutation targets in colorectal tumors with different degrees of microsatellite instability. Oncogene 1998;17:157-63.
189. Thibodeau SN, French AJ, Cunningham JM, Tester D, Burgart LJ, Roche PC, McDonnell SK, Schaid DJ, Vockley CW, Michels VV, Farr GH, Jr., O'Connell MJ. Microsatellite instability in

colorectal cancer: different mutator phenotypes and the principal involvement of hMLH1. Cancer Res 1998;58:1713-8.
190. Fodde R, Smits R, Clevers H. APC, signal transduction and genetic instability in colorectal cancer. Nat Rev Cancer 2001;1:55-67.

4. RESULTS (summary of the publications)

"Transcriptome profile of human colorectal adenomas" Jacob Sabates-Bellver*, Laurens G. van der Flier*, Mariagrazia de Palo, Elisa Cattaneo, Caroline Maake, Hubert Rehrauer, Endre Laczko, Michal A. Kurowski, Janusz M. Bujnicki, Mirco Menigatti, Judith Luz, Teresa V. Ranalli, Alfredo Pastorelli, Roberto Faggiani, Marcello Anti, Josef Jiricny, Hans Clevers and Giancarlo Marra. *In press, Molecular Cancer Research*. *These authors contribute equally to this work (**Appendix I**).

The objectives of this study were to characterize the global gene expression changes occurring during the transition from the normal to the adenomatous epithelium in the colon and to identify novel biomarkers of these tumors.

Using a standardized microarray platform covering the entire human genome, we analyzed the transcriptional features of 32 prospectively collected colorectal adenomas, each with matched samples of normal mucosa.

We first showed that the expression profiles of the adenomas could be easily distinguished from that of the normal mucosal samples by using different unsupervised analyses.

Gene Ontology analysis revealed that the biological processes more frequently represented by genes found up-regulated in adenomas were *mitosis*, cell division, DNA replication and spindle organization, whereas *host immune defense*, inorganic anion transport, organ development and inflammatory response were the most over-represented processes by genes found down-regulated in this lesions.

Since the Wnt signalling pathway is the physiological regulator of epithelial homeostasis, we performed a gene expression analysis of all the genes presently considered involved in this pathway. We found out that almost half of them (45%) showed statistically significant expression changes in adenomas, whereas 34% remained unchanged and 21% were expressed neither in the normal colorectal mucosa nor in the adenomas.

Using Canonical Correspondence Analysis (a supervised variant of the Correspondence Analysis), we correlated the gene expression profiles with the clinico-pathological characteristics of the adenomas. This analysis revealed that the expression profile of adenomas with a diameter >20 millimetres could be easily distinguished from that of smaller adenomas. This observation is relevant because the dimension of the adenomas is considered an prognostic factor. We are in the process of identifying the gene expression changes that are more relevant to the growth of adenomas.

We then focused on *KIAA1199*, a gene encoding a protein with an unknown function. This gene was over-expressed ~55 and ~39 fold in adenomas and carcinomas, respectively. We showed that *KIAA1199* expression level significantly correlate with that of many well known Wnt target genes. Furthermore, its expression was down-regulated after induction of dominant negative TCF4, beta-catenin siRNA, or N-terminal deleted TCF4 mutant, all of them inhibitors of the Wnt-pathway, suggesting its Wnt target status. These evidences confirmed the results obtained in our previous work ("The intestinal Wnt/TCF signature", Appendix II).

By *in situ* hybridization and immunohistochemistry we showed that KIAA1199 mRNA and protein were predominantly expressed in the proliferating compartment of normal intestinal crypts, with much higher expression levels in dysplastic and neoplastic glands. This expression pattern is reminiscent of that of many Wnt target genes. KIAA1199 was expressed in the cytoplasm, particularly in the luminal side of the dysplastic cell multilayer.

Based on our data, we concluded that KIAA1199 is a novel biomarker of colorectal neoplasia and displays potential as an interventional target.

"The intestinal Wnt/TCF signature" Laurens G. Van der Flier*, Jacob Sabates-Bellver*, Irma Oving, Andrea Haegebarth, Mariagrazia de Palo, Marcello Anti, Mareielle E. van Gijn, Saskia Suijkerbuijk, Marc Van de Wetering, Giancarlo Marra and Hans Clevers. *Gastroenterology 2007;* ***132****:628-32*. *These authors contribute equally to this work (**Appendix II**).

The goal of this publication was to describe a core target gene program, the Wnt/TCF signature gene set, responsible for the transformation of human intestinal epithelial cells.

Comparing the transcriptome of two colorectal cancer cell lines (stably transfected with dominant negative TCF4 or TCF1 to block the Wnt signalling pathway) with that of colorectal adenomas, adenocarcinomas and normal mucosa (were the Wnt signalling pathway is abnormally activated), we found out new target genes of this pathway.

We identified new putative Wnt target genes related to the different stages of carcinogenesis; adenomas, adenocarcinomas, and both tumor conditions (51, 36 and 121 genes, respectively).

By *in situ* hybridization on Apc^{min} mice (these mice carry a mutant *Apc* allele, developing multiple adenomas, primarily in the small intestine), and based on the staining patterns observed in the crypts, the putative Wnt target genes could be grouped in three different categories. Genes in the first category were expressed in the proliferative compartment of the crypts. Approximately half of the genes displayed a gradient of decreasing signal along the base of the villus. For the remaining genes in this category, mRNA expression was restricted to the proliferative compartment of the crypt proper with a steep gradient towards the crypt-villus junction. The second category comprised Paneth cell maturation markers. In the third category we found genes whose staining pattern is limited to 1-5 crypt cells near the crypt bottom, were the crypt stem cells have been mapped previously.

Basically, the Wnt/TCF signature gene set defines the core program activated by TCF4 in intestinal epithelial cells . The individual genes within the signature represent promising targets for therapy of colorectal cancer, since their expression is consistently activated as the direct result of oncogenic Wnt pathway mutations, while many target genes will be causally involved in the transformed behavior of the neoplastic cells.

"Defective DNA Mismatch Repair Determines a Characteristic Transcriptional Profile in Proximal Colon Cancers" Massimiliano di Pietro*, Jacob Sabates Bellver*, Mirco Menigatti, Fridolin Bannwart, Annelies Schnider, Anna Russell, Kaspar Truninger, Josef Jiricny and Giancarlo Marra. Gastroenterology 2005; 129:1047-1059. *These authors contribute equally to this work (**Appendix III**).

The aim of this study was to characterize the transcriptional profile of mismatch repair deficient (MMR-) colon carcinomas, which represent a sub-group of colon cancers with peculiar molecular, pathologic and clinical features (microsatellite instability, conspicous lymphocityc infiltration, preferential localization in the proximal colon, and better prognosis).

In an unsupervised analysis of microarray data, we showed that MMR status (MMR- or MMR+) was the most important variable which allowed an almost perfect clustering separation of the proximal colon cancers included in our series. We identified a *MMR-status signature* of 100 genes, whose gene expression changes perfectly segregated the MMR+ cancers from the MMR- deficient ones and that most likely explain the clinico-pathological differences between these two kinds of colon cancers.

We also identified gene expression alterations, relative to expression in normal colon mucosa, specific to mismatch repair deficient (MMR-) or mismatch repair proficient (MMR+) colon cancers. We named this group of 39 genes the *MMR-related tumorigenesis signature*.

In order to isolate the changes more likely localized in the epithelial-cell component of the colon cancers (avoiding a possible dilution effect due to presence of stroma) we compared the gene expression profiles of the cancers with those of the colon cancer epithelial cell lines. Genes with similar expression changes (MMR+ > MMR- or MMR- >MMR+) in the cancer samples and the colon cancer cell lines presumably reflect an epithelial-cell specific alteration. We designated this group of 54 genes as the *MMR-related epithelial-cell signature*.

Since genetic alterations caused by MMR deficiency itself (e.g., microsatellite instability within the coding or regulatory regions of genes expressed by cancerous epithelial cells) might lead to some of the transcriptional differences seen between MMR- and MMR+ cells, we analyzed the frequency of microsatellites in the cDNAs of the genes contained in the MMR-related epithelial-cell signature by comparing them to 100 control genes. We found out that microsatellites consisting of $\geq$8 repeat units in the coding region and/or in the 5'UTR and /or $\geq$12 repeat units in the 3'UTR were significantly more common and longer in the *MMR-related epithelial-signature* genes than those observed in control genes. We suspected that MSI in these microsatellites might affect the stability of the transcripts. We confirmed microsatellite instability in repeats of the 3'UTR of three genes found differently expressed in MMR- cancers: *NUTF2*, *HNRPL* and *CTNNB1*.

Due to the conspicous lymphocytic infiltration characteristic of MMR- cancers, we focused on genes involved in the immune response against tumor cells. By using real-time RT-PCR, we confirmed the high mRNA levels of *4-1BBL* (*TNSFS9*), a gene that plays a crucial role in the anti-tumor response, in MMR- cancers. Furthermore, the high expression of this gene was associated with a high expression of the corresponding protein in MMR-deficient colon cancer cell lines studied by flow cytometry. We proposed 4-1BBL as a biomarker of MMR-deficient colon cancers and suggest its

potential as a target for therapeutic drugs development supported by the fact that it is not expressed in the cells of the colon normal mucosa.

Finally, we compared the transcriptome of colon cancers (independently of their MMR status) with that of the colon normal mucosa to find out genes important for colon carcinogenesis in general. We find 182 genes with significantly different expression levels between normal mucosa and colon cancers with a fold change ≥4. The involment of many of these genes in colon carcinogenesis has been reported in previous studies, but a number of gene expression changes are novel findings.

It is worth mentioning that microarray data were obtained using the HG-U95v2 arrays from Affymetrix, which contain approximately 12000 probes. However, we have recently hybridized the same RNA samples with Affymetrix HG-U133plus2.0 arrays, which contain oligonucleotides representing the entire human genome. These new results will give us a much wider perspective on the gene expression changes in colon cancer (these data will be the subject of a different study).

"The transcriptome of ileal and colonic normal mucosa" Jacob Sabates-Bellver, Tobias Gonzenbach, Endre Laczko, Hubert Rehrauer, Caroline Maake, Fridolin Bannwart, Annelies Schnider, Josef Jiricny and Giancarlo Marra. Manuscript in preparation (**Appendix IV**).

The aim of this study was to obtain a comprehensive comparative transcriptional analysis of human colon and ileum normal mucosa tissues to find out genes and pathways responsible for the different incidence of adenocarcinomas (in colon is much higher than in ileum) in these two GI tracts considered so far rather similar in morphology and cellular organization.

Twelve paired colonic and ileal normal mucosa samples from 12 patients were used for microarray data analysis. Different unsupervised analyses of the transcriptional levels of the expressed genes effectively segregate the two groups.

We performed Gene Ontology analysis with the list of genes whose expression in the two parts of the GI tract differed significantly. We found that the genes more expressed in colon were primarily implicated in the positive regulation of cell motility and cell migration while those more expressed in ileum were principally associated to different aspects of metabolism, protein kinase C activation and regulation of apoptosis, among others.

Because several signalling pathways maintain the cellular homeostasis in the GI epithelium, we looked for changes in the expression levels of genes involved in these pathways, as well as of numerous transcription regulators.

The Wnt signalling pathway is the physiological regulator of epithelial homeostasis and has been described that the transformation process to adenoma, and lately to carcinoma, begins in the epithelial crypts as a result from qualitative, quantitative and spatial subversion of this pathway. For this reason, we also searched for differently expressed genes involved in this pathway.

Given that cell renewal rates in ileum and colon are very high and that apoptosis plays a key role in this process, we focused our attention on the apoptosis-related genes, particularly in the

expression levels of *CIDEB* and *CIDEC*, two members of the cell death-inducing DFFA-like effectors family. We have found a significant over-expression of these two genes in ileum compared with colon normal mucosa. Interestingly, *CIDEB* and *CIDEC* mRNA expression levels were further down-regulated in adenomas and adenocarcinomas of the colon. We are currently analyzing their expression at the protein level.

These preliminary results suggest that CIDEB and CIDEC may play an important role in colorectal tumorigenesis prevention.

"Expression of the MutL Homologue hMLH3 in Human Cells and its Role in DNA Mismatch Repair" Elda Cannavo, Giancarlo Marra, Jacob Sabates-Bellver, Mirco Menigatti, Steven M. Lipkin, Franziska Fisher, Petr Cejka and Josef Jiricny. Cancer Res 2005; 65 (23): 10759-66 (**Appendix V**).

In this paper, we showed the participation of *hMLH3* in MMR *in vitro* and its partial redundancy with *hPMS2*, which would be an explanation for the low penetrance of *hPMS2* mutations in hereditary nonpolyposis colon cancer families (HNPCC families).

A polyclonal antibody against the COOH terminus of hMLH3, which contains the hMLH1-interacting domain, was raised to avoid the inconvenient caused by the fact that the commercially available hMLH3 antibodies could detect the recombinant protein on WB but failed to detect the endogenous protein in all the human cell lines used in this study.

After having established that the level of hMLH3 in human cells is not dependent on hMLH1 and that correlates with *hMLH3* mRNA, we wondered whether the fluctuation of hMLH3 expression in the tested cell lines could be linked to the well known gene silencing process by cytosine methylation. The CpG island present in the promoter was found to be methylated giving the mechanism to explain, although this is not the only one, the lack of expression of hMLH3.

Given that hMLH3, hPMS2 and hPMS1 interact with the same region of hMLH1, the correlation between the relative abundance of the three different heterodimers and the phenotype of the human cells was assessed. We showed that hMLH3 is less abundant than hPMS1 and hPMS2.

We also showed how, differently from hPMS2 and hPMS1, the presence of hMLH1 is not required for the stabilization of hMLH3 in human cells, although it is necessary in baculovirus-infected Sf9 cells.

Finally, we speculated that physiologic levels of hMutLγ might not be sufficient to mediate mismatch correction *in vivo* although in our *in vitro* MMR assays, the factor could participate in the correction of base-base mispairs and one-nucleotide insertion/deletion loops.

My direct contribution to this article was concentrated at the RNA level of hMLH3 in cell lines, producing the microarray data. We could show a correlation between the RNA and the protein levels of hMLH3 in the different cells used in this study.

5. Conclusions and perspectives

This section is divided in four different parts, each of them related to one of the projects of my thesis (see results and appendixes).

Transcriptome profile of human colorectal adenomas

Our data represents the first comprehensive transcriptomic analysis of these precancerous lesions, and provide evidences of the existence of a broad-scale transcriptome remodeling during adenoma formation.

Expression data subjected to four different unsupervised analyses - Hierarchical clustering, Principal component analysis, Correlation analysis, and Correspondence analysis- effectively segregated the normal tissues from the adenomas.

Furthermore, the expression profile of adenomas measuring > 20 mm was easily distinguished from those of smaller (= 20 or < 20 mm) adenomas. As expected, larger diameters were correlated with high-degree dysplasia.

We have focused on a series of changes we consider crucial in adenomatous transformation in particular those involved in i) regulation of transcription, ii) cell proliferation, cell differentiation and apoptosis, iii) cell adhesion, and iv) the Wnt signaling pathway.

We have identified *KIAA1199* as a novel target of the Wnt signaling pathway based on several evidences: i) its protein expression at the bottom of the intestinal crypts where the proliferative compartment is placed, ii) its up-regulation in adenomas and carcinomas compared to normal mucosa, and iii) its down-regulation after blocking the beta-catenin/TCF transcription complex in CRC cell lines. Due to the over-expression in colon tumors and to its membrane's surface localization, we also proposed *KIAA1199* as a new molecular biomarker of colorectal tumors with capabilities of being a potential interventional target.

Because the function of KIAA1199 is still unknown, our next objectives are to find out i) the regulation mechanisms of its expression, ii) the morphological and molecular changes triggered by its up- or down-regulation in cell lines, iii) the phenotype of KIAA1199 knock-out animal models, iv) whether its up-regulation is specifically linked to colorectal tumors or common to other cancerous lesions.

We will also analyze transcriptional changes between colorectal adenomas and carcinomas in order to track expression changes along the entire transformation process from normal mucosa to invasive cancer.

The intestinal Wnt signaling pathway

To identify new Wnt target genes is a subject of great interest to better understand cancer progression. On this regard, we have performed a combined transcriptomic analysis using colorectal cancer cell lines, normal mucosa, adenoma, and carcinoma colon samples, to identify new targets of this pathway. With this approach (see Appendix II), we have found putative Wnt target genes related

to the different progression stages of the carcinogenic process. We classified these genes, based on their expression localization, in three different groups: i) genes expressed in the proliferative compartment of the crypts. Approximately half of the genes displayed a gradient of decreasing signal along the base of the villus. For the remaining genes in this group, mRNA expression was restricted to the proliferative compartment of the crypt proper with a steep gradient towards the crypt-villus junction, ii) Paneth cells maturation markers, and iii) putative stem cells markers, i.e., genes whose staining pattern was limited to 1-5 crypt cells near the crypt bottom.

At this point, as we did for *KIAA1199*, these putative Wnt target genes should be individually validated with other techniques (such as Northern blotting, *in situ* hybridization, reporter gene experiments, ChIP on chip, immunohistochemistry, etc) to prove that their expression is certainly regulated by this pathway.

Transcriptional profile of mismatch repair deficient colon cancer

In order to find out molecular alterations related to the MMR status, we conducted a gene expression profile analysis by comparing the transcriptome of MMR-deficient with that of the MMR-proficient tumors and cell lines.

We identified a *MMR-status signature* of 100 genes, whose gene expression changes perfectly segregated the group of MMR+ cancers from that of MMR- deficient.

Gene expression alterations specific to mismatch repair deficient (MMR-) or mismatch repair proficient (MMR+) colon cancers, relative to expression in normal colon mucosa, were also identified and named as *MMR-related tumorigenesis signature*.

A *MMR-related epithelial-cell signature* containing epithelial-cell specific alterations was as well identified.

Our data showed several changes that might explain the better prognosis associated with MMR-deficient colon cancers. Compared with MMR-proficient cancers, MMR-deficient tumors presented reduced transcript levels of several genes with functions that presumably favor tumor progression and increased expression of factors believed to inhibit tumor growth, progression, and/or invasiveness.

Among the genes found to be more expressed in MMR-deficient tumors, we identified *4-1BBL* (*TNSFS9*), a key player in the anti-tumoral response mechanism, as a biomarker of this type of cancers. Given that *4-1BBL* is not expressed in the colon normal mucosa, its over-expression in MMR-deficient lesions could be exploited for a novel therapeutic approach in this subgroup of colon cancers.

We suspected that some of the expression changes might conceivably reflect transcriptional effects of MSI within the coding or regulatory regions of genes expressed by cancerous epithelial cells, and we confirmed such hypothesis by the finding of a higher frequency of microsatellites in the complementary DNAs of the MMR-related epithelial cell signature genes compared with that of randomly selected expressed.

The microarray platform we used screened only ~10000 genes. We have recently repeated the analysis using an array that covers the entire human genome. These new data confirmed our previous results and brought to light other significant gene expression changes. We are currently

analyzing the list of genes obtained to select interesting candidates for further investigation. We are also planning to combine these data with the information obtained from a brand new type of array, called Exon Array, which reveals alternative splicing of the genes.

The transcriptome of ileal and colonic normal mucosa

Although morphology and cellular organization in different tracts of the lower GI tract are considered quite similar, physiology and cancer incidence are strikingly different. Thus, we performed a transcriptome analysis in ileum and colon normal mucosa tissues in order to identify gene expression patterns that might offer new work hypotheses to achieve a better molecular understanding of these differences.

Amongst the genes significantly changed between the two groups, we focused on those involved in transcription regulation, cell proliferation/differentiation, Wnt signaling pathway, and apoptosis.

In the apoptosis-related list of genes, we found the cell death-inducing DFFA-like effectors *CIDEB* and *CIDEC* markedly more expressed in ileum than in colon. In addition, these two genes were further down-regulated in adenomas and adenocarcinomas of the colon. We are currently checking whether the CIDEB and CIDEC proteins are also differentially expressed in the two epithelia and how their expression is distributed along the crypt-villous axis. Based on these experiments, we will evaluate the feasibility of a new project aiming to verify whether the modulation of CIDEB and CIDEC expression in mouse ileum and colon might modify the incidence of adenocacinoma in these two tissues.

The analysis of genes involved in signaling pathways controlling the cellular homeostasis in the intestinal mucosa, such as Notch, TGF-beta, BMP, Hedgehog, or RAS/RAF, is also in progress.

6. Appendixes

6.1. Appendix I

“Transcriptome profile of human colorectal adenomas” Jacob Sabates-Bellver*, Laurens G. van der Flier*, Mariagrazia de Palo, Elisa Cattaneo, Caroline Maake, Hubert Rehrauer, Endre Laczko, Michal A. Kurowski, Janusz M. Bujnicki, Mirco Menigatti, Judith Luz, Teresa V. Ranalli, Alfredo Pastorelli, Roberto Faggiani, Marcello Anti, Josef Jiricny, Hans Clevers and Giancarlo Marra. Manuscript submitted for publication (October 2006). *These authors contribute equally to this work

Transcriptome profile of human colorectal adenomas

Running title: The transcriptome of colorectal adenomas

Jacob Sabates-Bellver*[#], Laurens G.Van der Flier[‡#], Mariagrazia de Palo^, Elisa Cattaneo*, Caroline Maake[§], Hubert Rehrauer[ʃ], Endre Laczko[ʃ], Michal A. Kurowski[¶], Janusz M. Bujnicki[¶], Mirco Menigatti°, Judith Luz[†], Teresa V. Ranalli[Δ], Vito Gomes[Δ], Alfredo Pastorelli^, Roberto Faggiani^, Marcello Anti^, Josef Jiricny*, Hans Clevers[‡], and Giancarlo Marra*[¥]

* Institute of Molecular Cancer Research, [§] Institute of Anatomy and [ʃ] Functional Genomic Center, University of Zurich, 8057 Zurich, Switzerland

[‡] Hubrecht Institute, Netherlands Institute for Developmental Biology, 3584CT, Utrecht, The Netherlands

^ Gastroenterology Unit and [Δ] Pathology Department, Belcolle Hospital, 01100 Viterbo, Italy

° Institute of Biochemistry and Genetics, and [†] Division of Medical Genetics, University of Basel, 4058 Basel, Switzerland

[¶] Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology, ul. Ks. Trojdena 4, 02-109 Warsaw, Poland.

[#] J.S-B and L.G.VdF. contributed equally to this study

[¥] Corresponding author:
Giancarlo Marra, M.D., Ph.D.
Institute of Molecular Cancer Research, University of Zurich
Winterthurerstrasse 190, 8057 Zurich, Switzerland
Tel. 0041 044 635 3472, Fax. 0041 044 635 3484, marra@imcr.uzh.ch

Keywords: colorectal adenomas; transcriptome; Wnt signaling; KIAA1199.

Abstract

Background: Colorectal cancers are believed to arise predominantly from adenomas. Although these precancerous lesions have been subjected to extensive clinical, pathological, and molecular analyses, little is currently known about the global gene expression changes accompanying their formation.

Results: To characterize the molecular processes underlying the transformation of normal colonic epithelium, we compared the transcriptomes of 32 prospectively collected adenomas with those of normal mucosa from the same individuals. Important differences emerged not only between the expression profiles of normal and adenomatous tissues, but also between those of small and large adenomas. A key feature of the transformation process was the remodeling of the Wnt pathway reflected in patent over- and underexpression of 78 known components of this signaling cascade. The expression of 19 Wnt targets was closely correlated with clear upregulation of *KIAA1199,* whose function is currently unknown. In normal mucosa, KIAA1199 expression was confined to cells in the lower portion of intestinal crypts, where Wnt signaling is physiologically active, but it was markedly increased in all adenomas, where it was expressed in most of the epithelial cells, and in colon cancer cell lines, it was markedly reduced by inactivation of the β-catenin/TCF(s) transcription complex, the pivotal mediator of Wnt signaling.

Conclusions: Our transcriptomic profiles of normal colonic mucosa and colorectal adenomas shed new light on the early stages of colorectal tumorigenesis and

identified *KIAA1199* as a novel target of the Wnt signaling pathway and a prominent early marker of colorectal adenomatous transformation.

3

Background

In developed countries, sporadic adenomatous colorectal polyps are found in roughly one third of asymptomatic adults below the age of 50 who undergo colonoscopy. Depending on their characteristics (multiplicity, size, histological features, degree of dysplasia), these lesions can be associated with a substantial risk of recurrence (up to 60% at three years) and the development of advanced neoplastic disease [1](and additional references herein). It has been estimated that 15% of all adenomas measuring 1 cm or more will progress to carcinomas within 10 years of their detection [2].

Although adenomatous polyps are not the only precancerous lesions in the colorectum, they are the most common, and they are the precursors of most of the cancers in this organ. In these neoplasms, the transformation process begins in the epithelial crypts and seems to result from qualitative, quantitative, and spatial subversion of the Wnt-signaling pathway, the physiological regulator of epithelial homeostasis [3-5]. This adenoma-carcinoma pathway of tumorigenesis is characterized by mutations involving various components of this pathway (e.g., *APC*, whose germline mutations are responsible for familial adenomatous polyposis, a cancer-predisposition syndrome that accounts for ~0.5% of colorectal cancers [6]; *CTNNB1,* which encodes a subunit of the cadherin protein complex known as β-catenin; and *Axin*, the gene encoding a multidomain scaffold protein that is essential for β-catenin degradation). The result of these mutations is an accumulation of β-catenin, first in the cytoplasm and then in the nucleus, where it associates with DNA-

binding proteins of the T-cell factor (TCF)/lymphoid-enhancer factor (LEF1) family, transforming them from transcriptional repressors into transcriptional activators that affect the expression of numerous genes involved in epithelial homeostasis.

Although the key role played by adenomatous polyps in colorectal tumorigenesis is widely acknowledged, the gene expression changes that trigger or accompany their development have never been comprehensively studied. We therefore conducted a transcriptomic analysis of prospectively collected colorectal adenomas, using a standardized oligonucleotide microarray covering the entire human genome. This study not only provided new information that is fundamental for future molecular characterization of these precancerous lesions, it also allowed us to identify a novel biomarker of colorectal tumorigenesis.

Results

The focus of our study was the adenoma-adenocarcinoma pathway of colorectal carcinogenesis, which is closely linked to deregulation of the Wnt signaling pathway. To gain insight into the early steps of this process, we confined our investigation exclusively to sporadic (i.e., non-familial), pedunculated colorectal adenomas (type *0-Ip* of the Paris classification [7]). Nonpolypoid and sessile polypoid lesions were not included because in some cases their transformation is believed to proceed along non-adenomatous pathways [8]. Details on our case selection criteria are provided in Materials and Methods.

Thirty-two pedunculated adenomatous polyps, each with matched samples of normal mucosa, were prospectively collected from 28 patients (**Table 1**). The total number of synchronous and previously excised adenomas was < 3 in 18/28 patients and 3-15 in the remaining ten. In this latter subgroup, the absence of *APC*- or *MYH*-associated multiple adenomatosis had been confirmed by genetic testing. Histological analysis of one polyp (case NM) revealed superficial infiltration of the submucosa, but this case was not excluded since the region sampled for microarray analysis was clearly adenomatous. (As noted below, this finding was consistent with the results of hierarchical cluster analysis shown in Additional data file 1.)

Analysis of microarray data for the 32 adenoma/normal mucosa tissue pairs revealed that 31,033 of the probes were expressed in one or both of the tissue groups. The normal tissues were effectively segregated from the adenomas in four unsupervised analyses of the expression levels of these genes [hierarchical clustering, principal component analysis (PCA), correlation analysis and correspondence analysis (CA); see Materials and Methods for details] (**Figure 1**). In a separate analysis, these two tissue groups were also unequivocally distinguished from a previously described set of 25 colon cancers [9], which we reanalyzed for this study with the same microarray used to characterize the adenomas and normal mucosa (**Additional data file 1**).

Almost half of the expressed probes (15,059 / 31,033) displayed significant expression changes in adenomas. This finding was not unexpected since the two tissue types are characterized by dramatic histological differences. Some of the

observed changes undoubtedly reflect the different proportions of epithelial and stromal cells in the two tissues rather than real transcriptional changes in the neoplastic epithelium. However, we devised a sampling procedure (see Materials and Methods for details) that allowed us to obtain samples containing a high percentage of epithelial-cell RNA of high quality and in amounts large enough to require only a single round of amplification. As a result, we did not need to resort to microdissection, which in our experience considerably reduces the quality of the extracted RNA. The complete microarray data sets are available (one of the public repositories). Our analysis focused almost exclusively on expression differences that were $\geq$ 2-fold since less marked changes were more likely to be non-relevant to the pathogenesis of adenomas.

Probes with fold changes $\geq$ 2 (1190 probes upregulated and 2469 downregulated in adenomas) were subjected to gene ontology analysis to identify the biological processes involved in the transition of normal mucosa to adenoma. The most significant results of this analysis are listed in **Additional data file 2.** The processes that were most markedly over-represented among genes that were upregulated in adenomas included *mitosis*, *DNA replication* and *spindle organization.* Downregulated genes were predominantly involved in *host immune defense*, *inorganic anion transport*, *organ development*, and *inflammatory response*, even though a small group of genes involved in the latter process were upregulated in adenomas (**Additional data file 3).**

7

We then analyzed the transcript levels of 319 genes believed to be components of the complex Wnt signaling pathway (**Additional data file 4**). Sixty-six of these genes (21%) were not expressed in either the normal or adenomatous tissue, and 34% were expressed similarly in both tissue groups. The remaining 144 genes - almost half (45%) of those linked to the Wnt pathway – displayed significantly altered expression in adenomas, and 78 / 144 displayed fold changes of $\geq$ 2.

A supervised extension of CA [10], canonical correspondence analysis (CCA), was then used to identify possible correlations between gene expression patterns and clinical or pathological variables. Four of the variables considered (*adenoma diameter, colon segment of origin, degree of dysplasia,* and *adenoma recurrence* - see Table 1) were clearly associated with distinct clusters of expression profiles (**Figure 2, panel A**). The profile of adenomas measuring > 20 mm could be easily distinguished from those of smaller (= 20 or < 20 mm) adenomas (**Figure 2B**). It is interesting to note that *adenoma diameter* was closely correlated with (i.e., its vector in panel A is almost parallel to) CCA axis 1, which represents the most important proportion of the total variance. Larger diameters were correlated with high-degree dysplasia, as would be expected.

The CCA plot of the 11,709 modeled probes (loading plot, not shown) suggested that the distinction between the three size groups of adenomas is due to a complex network of relatively small changes in the expression of numerous genes (as opposed to marked changes involving a limited number of genes). Nevertheless, to maximize the use of the extensive data sets, we selected the 500 probes with the

highest loading scores along the CCA axis 1 and isolated a set of genes whose expression changes displayed significant positive or negative correlation with adenoma size (**Additional data file 5**). Although their association with adenomas must be validated in a larger series, these are the expression changes most likely to play causal roles in the progression of these tumors.

It should be mentioned that normal mucosa from the sigmoid colon had an expression profile that differed significantly from that of tissues from other colon segments (**Figure 2A**). This finding will be explored in a future study conducted on a large series of normal mucosa samples from different colorectal segments.

The transcriptional profile of the 32 adenomas was thoroughly analyzed to identify genes likely to be involved in the development and evolution of these lesions. One of the first features that attracted our attention was the marked upregulation of *KIAA1199*, a gene encoding a protein with unknown function. Its overexpression was striking in all 57 colorectal tumors we examined (average increases of 54.8-fold in adenomas and 38.8-fold in carcinomas [9], compared with normal mucosa). These findings were fully confirmed by real-time RT-PCR analysis of RNA extracted from samples used for the microarray study and from additional samples collected after the present study was completed (**Additional data file 6**).

In light of these findings, it was natural to wonder whether *KIAA1199* might be a novel positively regulated target of Wnt signaling, which is characteristically deregulated in colorectal tumors. Previous microarray studies indicated that genes co-regulated at the transcriptional level under different conditions tend to be involved

in the same processes and pathways, and the analysis of transcriptional co-expression has been used to predict the function of novel genes [11-13]. Therefore, we conducted a search for known Wnt targets (listed in **Additional data file 7**) among the genes whose expression patterns in all the tissue samples significantly correlated with those of *KIAA1199.* (The procedure used in this analysis is summarized in Materials and Methods and **Additional data file 8**.) Forty-nine percent of the known Wnt targets that were overexpressed in our adenoma samples had expression patterns that were positively correlated with that of *KIAA1199* (**Figure 3, panels A and B**), as opposed to only 7.9% of the overexpressed genes that are not considered Wnt targets ($p<0.0001$).

Evidence of the potential involvement of *KIAA1199* in the Wnt signaling pathway had also emerged from another study by our group (Van der Flier et al., *Gastroenterology*, in press). A combined analysis of microarray data of tissues and cell lines placed *KIAA1199* at the top of a list of genes (Supplementary Table 1 of Van der Flier et al.) that were upregulated in colorectal adenomas and downregulated in colon cancer cell lines that had undergone stable transfection with doxycycline-inducible forms of dominant negative TCF(s) (dnTCF1 or dnTCF4) to suppress Wnt signaling [14, 15]. In the present study, *KIAA1199* was also found to be markedly downregulated in LS174T colon cancer cells in which Wnt signaling had been blocked by the induction of β-catenin siRNA or N-terminal deleted TCF4 [15, 16]. The dramatic decrease in *KIAA1199* mRNA levels associated with this inhibition of the Wnt pathway was confirmed by Northern blotting (**Figure 3C**).

In general, Wnt target genes are expressed predominantly in the proliferating compartment of normal intestinal crypts (lower portion), and their expression is appreciably increased in adenomatous glands [15]. Our analysis of human tissues with preserved architecture indicated that these are also attributes of *KIAA1199*. In *in situ* hybridization studies, *KIAA1199* mRNA was detectable only in the lower portion of normal colonic epithelial crypts (**Figure 4A and B**), and its expression levels were much higher in dysplastic glands (**Figure 4C**). These patterns were confirmed at the protein level by immunohistochemistry performed with an antibody raised in our laboratory (**Figure 4D-K**). This analysis also revealed that the KIAA1199 is a cytoplasmic protein whose expression is most intense near the cell membrane, particularly on the luminal side of the dysplastic cell multilayer (**Figure 4F-K**).

Discussion

Adenomatous colorectal polyps are one of the most common human tumors and the most frequent precancerous lesions in the colorectum, but their transcriptome has been only partially analyzed, and the data are generally based on a limited number of cases [17-19]. We attempted to fill this gap by performing a comprehensive whole-genome microarray analysis of a large, highly homogenous set of adenomas that was collected prospectively.

A comparison of the transcriptomes of adenomatous polyps and segment-matched samples of normal colorectal mucosa revealed evidence of broad-scale

remodeling. As a starting point for future verification studies, we have drawn up a list of 478 genes that were significantly up- (n=153) or downregulated (n=325) in the adenomatous tissues (fold changes of 4 or more) (**Additional data file 9**). Space constraints preclude more than a cursory examination of this list, but we have highlighted in **Table 2** certain aspects that we feel are particularly interesting in terms of their relevance to the process of adenoma formation. For instance, transcription regulation appears to be extensively modified. Twenty-nine molecules involved in this process were expressed in adenomas at levels > 4 higher or lower than those observed in the normal mucosa, but there were also a number of smaller changes in this category (**Additional data file 10)** that might also have dramatic effects on gene expression. Several other alterations are noteworthy in terms of their potential impact on cell proliferation, differentiation, apoptosis, and cell adhesion: 1) Upregulation of four members of the REG (*Reg*enerating) family of genes [20, 21], which would lead to increased tissue-mitogen expression; 2) Upregulation of *LCN2* [22] and downregulation of *ZFHX1B/SIP-1* [23] in the absence of significant changes in the expression of the epithelial cadherin *CDH1* (E-cadherin), which would prevent or delay the epithelial-mesenchymal transition. (Changes were also noted in the expression of other cell-adhesion genes of the cadherin and claudin families, including the striking overexpression of the placental cadherin gene *CDH3*, which is associated with early events in the transformation process [24, 25]); 3) Downregulation of *ZFHX1B/SIP-1* and *Max dimerization protein 1* (*MXD1/MAD1*) (decreased only 3.3 fold and therefore not listed in Table 2) [26, 27] and

overexpression of the *RTEL1* helicase, which should facilitate telomere elongation [28]; 4) Alterations that would diminish apoptosis, e.g., overexpression of the decoy receptor for Fas ligand, *TNFRSF6B*, which is reportedly co-regulated with *RTEL1* on chromosome 20q13.3 [29-31]; decreased expression of the netrin-1 receptor, *UNC5C* [32]; and expression changes involving three Fas apoptosis inhibitory molecules (*FAIM*), including *FAIM1*, which was increased 2.3-fold and is thus not listed in Table 2; 5) Marked downregulation of several genes that would result in reduced tumor-suppression activity, e.g., those encoding the anti-angiogenic factor ANGPTL1 [33], the cyclin-dependent kinase inhibitor CDKN2B/p15, and the prostaglandin catabolism enzyme HPGD [34]. It is also important to recall the size-related differences noted in the adenoma gene expression profiles (Figure 2 and Additional data file 5). When validated in a larger series of tumors, these differences should provide important clues to the molecular basis of the well known link between the dimensions and malignant potential of colorectal adenomas [1].

Our study also furnishes a complete picture of expression changes involving gene components of the Wnt pathway across the transition from normal to adenomatous epithelium (Additional data file 4), as well as evidence for the existence of a novel Wnt target: *KIAA1199.* This gene, which encodes a protein of unknown function, was strikingly overexpressed in all the adenomas included in this study and in 25 adenocarcinomas of the colon described in a previous report [9]. Even more intriguingly, its expression was significantly correlated with that of several genes that are well established targets of Wnt signaling. Our hypothesis that *KIAA1199* is

upregulated by the TCF(s)/β-catenin transcription complex was considerably strengthened by the marked decreases in *KIAA1199* expression observed in cultured colorectal cancer cells when the Wnt pathway was inhibited by overexpression of dominant negative TCF4 proteins or by β-catenin knock-down. It is not yet clear whether this is a direct effect, but this possibility is supported by the results of a recent genome-wide TCF4 ChIP-on-chip analysis, which indicates that the *KIAA1199* locus is surrounded by four TCF4-bound regions (Hatzis, van der Flier and Clevers, unpublished).

Other features of *KIAA1199* expression are also compatible with its putative role as a Wnt target gene. *KIAA1199* mRNA and protein are both confined to the proliferative compartment of normal intestinal crypts, where Wnt signaling is normally active, and they are highly overexpressed in colorectal adenomas and carcinomas, where this pathway is almost always aberrantly activated.

In normal and tumor tissues, KIAA1199 is expressed in the cytoplasm of epithelial cells. In glands with low-degree dysplasia, higher concentrations are observed in the mucin vacuoles of goblet cells, but cytoplasmic expression of the protein in tumor cells remains elevated even after goblet-cell differentiation has been lost (Figure 4). These features, together with KIAA1199's localization in the luminal portion of the cytoplasm, are suggestive of a secreted and/or membrane protein. This conclusion is consistent with our *in silico* analysis of KIAA1199 (see **Additional data files 11 and 12**), which strongly predicts the presence of a signal peptide at its N-terminal end. In addition, the central region of KIAA1199 contains a TMEM2-

homology domain, which is present in a number of eukaryotic proteins, including TMEM2, polyductin (PKHD1), and fibrocystin L (PKHD1L1) (**Figure 5**), all large receptor proteins characterized by an N-terminal signal peptide or a single trans-membrane helix and a short cytoplasmic tail [35].

A study based on yeast two-hybrid screens suggested that KIAA1199 may interact with plexin A2 (KIAA0463) [36]. The trans-membrane plexins interact with trans-membrane semaphorins on nearby cells, providing "stop" and "go" signals that are crucial for cell motility and invasive growth [37, 38]. KIAA1199/plexin A2 interaction could thus play important roles in colorectal tumorigenesis, not only in the invasive stages, but also earlier during the formation of abnormal glands in benign adenomas.

A recent report linked high levels of *KIAA1199* mRNA with cell mortality in human fibroblasts and in a renal-cell carcinoma cell line [39]. In that study, however, there was no significant increase in *KIAA1199* expression during replicative aging of mortal cells, and this finding contrasts with the documented behavior of other genes involved in cell aging [40]. Furthermore, the authors reported wide variation in *KIAA1199* mRNA expression in breast cancer cell lines, and this finding raises the possibility that expression of this gene *in vivo* and in cell lines may differ.

Conclusions

We believe that our microarray data will serve as a springboard and reference point for other studies on the molecular basis of colorectal transformation

along the adenoma-carcinoma pathway (and subsequently for the study of alternative pathways as well). Some of the transcriptional changes reported in this study might one day be used as molecular indices of the susceptibility of adenomas to malignant transformation, information that would be helpful in planning appropriate follow-up of the lesions. As for KIAA1199, its invariably high expression in the colorectal tumors we studied raises interesting possibilities for the development of a new molecular biomarker for the diagnosis of these neoplasms. For example, since KIAA1199 expression in the normal mucosa is limited to cells in the lower portion of the crypts, which are not yet programmed to be shed into the intestinal lumen, the presence of KIAA1199 peptides in fecal water might prove to be a specific marker of adenomatous lesions. In addition, although due consideration must be given to its probable physiological role(s) in intestinal crypts and possibly in a number of other human tissues [39, 41, 42]), KIAA1199 may be a potential target of antibody-based therapies.

Materials and Methods

Tumor samples

Pedunculated colorectal polyps and normal mucosa were obtained during colonoscopies carried out in the Gastroenterology Unit of the Belcolle City Hospital in Viterbo, Italy. The tissues were collected prospectively with informed patient consent and the approval of the local Human Research Ethics Committee. Patients with documented familial polyposis, with > 15 adenomatous polyps (total: synchronous +

previously excised) [6], or currently treated with non-steroidal anti-inflammatory drugs (including aspirin) were excluded from the study.

For each polyp, three biopsies of normal mucosa were collected from the same colon segment (≥ 2 cm from the site of the polyp). Immediately after removal, a small sample of epithelial tissue (5-15 mg) was cut from the tip of each polyp, leaving the underlying *muscularis mucosae* intact. We excluded polyps < 1 cm to ensure that the sampling procedure would not interfere with the histologic diagnosis. All polyp samples were collected by a single operator (MdP) using the same procedure to minimize artifacts due to sampling differences. The approach used allowed us to obtain specimens with a high percentage of epithelial cells without resorting to microdissection, which can diminish the quantity and quality of the extracted RNA.

The polyp sample and the three normal mucosal biopsies were immersed in *RNAlater* (Ambion, Huntington, UK) for subsequent microarray analysis, and the remainder of the polyp was submitted for pathological analysis. The cut surface at the tip was labeled with India ink so that the sampled area could be easily identified during routine histologic examination. The tissue was then fixed in buffered formalin and embedded in paraffin. DNA extracted from sections of this specimen was also used to rule out microsatellite instability (reflecting defective DNA mismatch repair) at the *BAT26* locus, as previously described [43]. All of the polyps included in the study met the following criteria: type *0-Ip* [7]; maximum diameter: 1-4 cm; absence

of surface ulceration; histological diagnosis of adenoma; absence of microsatellite instability at *BAT26*.

In some analyses, we also included transcriptomic data from a previously described set of 25 colon cancers (mismatch repair-proficient and -deficient) [9], which we re-analyzed for this study with the same microarray used to characterize the adenomas and normal mucosa.

Microarray analysis, real-time RT-PCR, and Northern blotting

Total RNA was extracted (RNeasy Mini kit, Qiagen) from homogenized tissue samples (5- to 15-mg), and its integrity was verified by capillary gel electrophoresis (Bio Analyzer, Agilent Technologies). Complementary RNA (15 µg / sample), synthesized and labeled as previously described [9, 44], was hybridized with the Affymetrix U133 Plus 2.0 array (High Wycombe, UK), which contains *in situ* synthesized oligonucleotides representing the entire human genome (54,675 probes).

Raw gene-expression data generated by GCOS software (Affymetrix) were imported into the GeneSpring software program (Agilent) and normalized per chip (i.e., to the median of all values on a given array) and per gene (i.e., to the median expression level of the given gene across all samples). Analysis was performed using the log expression values with GeneSpring's cross-gene error model turned on. Probes were excluded from analysis unless they were listed as "present or marginal calls" and/or had expression values ≥100 in 50% or more (≥16 / 32) of the samples in at least one of the tissue groups (adenomas, normal mucosa).

Expression data were subjected to four different unsupervised analyses: 1) Hierarchical clustering using the Pearson correlation coefficient as a similarity measure and the average linkage algorithm for branch merging; 2) Principal component analysis (PCA), which reduces the dimensionality (number of variables) of a data set while retaining most of its variance [9]; 3) Correlation analysis, which involved computation of Pearson correlation coefficients for all possible sample pairs and visualization of correlation values as tile plots; 4) Correspondence analysis (CA), another dimension-reducing method [45], that was used to identify samples associated with particular gene expression levels. In typical CA, a matrix of *n* gene-expression levels from *p* samples is treated as a two-way contingency table (genes by samples or vice versa) with *n* and *p* specifications for the "factors" gene and sample, respectively. Each intensity value thus reflects the abundance of a given transcript in a given sample. Like PCA, CA identifies independent "factorial components" that account for variance within a multidimensional gene data set, but in this case the components are identified and ranked according to the correlation between gene and sample scores. A supervised or constrained extension of CA [10], canonical correspondence analysis (CCA), was then used to identify possible correlations between gene expression patterns and clinical or pathological variables. CA and CCA, as well as the corresponding plots, were computed using R software and the ade4 and made4 packages furnished by Bioconductor (http://www.bioconductor.org.)

19

The Mann-Whitney test was used to select genes differentially expressed in normal mucosa and adenomas; Benjamini-Hochberg multiple testing correction was applied with a false-discovery rate of 0.01. The genes in this set that were differentially expressed with fold differences of $\geq$ 2.0 were then analyzed with ErmineJ software [46] to identify any biological processes from the Gene Ontology database [47] that were over-represented.

Pearson correlation was used to identify correlation between *KIAA1199* expression and the expression of other genes in the entire set of tissue samples. Fisher's exact test was used to identify possible over-representation of known Wnt targets among genes whose expression was closely correlated with that of *KIAA1199* (correlation values ≥0.8).

RT-PCR and Northern blotting were performed as previously described [44, 48] to verify the expression level of *KIAA1199* in tissue samples and in LS174T colon cancer cells in which inducible inhibition of the Wnt pathway had been achieved with previously described methods [14-16].

In situ hybridization

Digoxigenin-labeled *KIAA1199* antisense riboprobes were synthesized from a PCR product amplified from human colon cDNA with *KIAA1199*-specific primers (sense: 5'cacatcggggaggagataga3'; antisense, containing a T7 RNA polymerase-binding site: 5'taatacgactcactatagggttccagacttgaca3'). This product was transcribed *in vitro* using the DIG RNA labeling kit and T7 RNA polymerase (Roche Diagnostics, Rotkreuz, Switzerland). *In-situ* hybridizations were performed on paraffin-embedded

sections of human colon fixed with 4% buffered formalin as described elsewhere [49].

Immunohistochemistry

Our *in silico* analysis of KIAA1199 (see Additional data file 11) indicated that residues 202-217 (IHSDRFDTYRSKKESE) form a loop between a conserved β-strand and the following helix of the N-terminal GG domain. This charged, surface-exposed peptide was used to raise a rabbit polyclonal antibody, which was purified by affinity chromatography on Thiopropyl Sepharose 6B (Amersham, Piscataway, NJ) derivatized with the antigenic peptide. A 1:1000 dilution of this antibody was used, as previously described [43], to evaluate KIAA1199 expression in formalin-fixed, paraffin-embedded sections of adenoma and normal mucosal tissues.

Abbreviations: PCA: Principal component analysis; CA: Correspondence analysis; CCA: Canonical correspondence analysis;

Author contributions

J.S-B. performed microarray analyses, antibody production, and other experiments as part of his PhD thesis; L.G. VdF. performed studies with cell lines and Northern blotting; M.dP. performed tissue sampling in adenomas and collected clinical data as part of her internship in Gastroenterology; E.C. performed extraction of nucleic acids and immunohistochemistry; C.M. performed in situ hybridizations; H.R performed image and data analysis and most of the statistical analyses; E.L. performed

correspondence and canonical correspondence analyses; M.A.K. and J.M.B performed *in silico* analysis of the KIAA1199 protein; M.M. and J.J. extracted nuclei acids, performed microsatellite analysis, and contributed to real-time RT-PCR experiments; T.V.R and V.G. independently analyzed and histologically classified tumor samples; A.P. and R.F. performed endoscopies with polypectomy and patient follow-up; M.A. enrolled patients, directed the clinical part of the study, and served as supervising mentor for M.dP.; J.J. and H.C. conceived and designed important experiments during the study and contributed to the writing of the manuscript; G.M. conceived the project, prepared the manuscript, and served as supervising mentor for J. S-B. during his PhD. Funding for this project came from the Zurich Cancer League and the Swiss National Science Foundation.

Acknowledgments:

We are grateful to the patients who participated in this study; to R. Haider, E. Pani, T. Lehmann, A. Patrignani, and U. Wagner for technical assistance; to A. Mueller, E. Veljkovic, and F. Bannwart for critical reading of the manuscript; and to M. Kent for editorial assistance. This study was supported by the Zurich Cancer League and the Swiss National Science Foundation.

References

1. Winawer SJ, Zauber AG, Fletcher RH, Stillman JS, O'Brien MJ, Levin B, Smith RA, Lieberman DA, Burt RW, Levin TR *et al*: **Guidelines for colonoscopy surveillance after polypectomy: a consensus update**

by the US Multi-Society Task Force on Colorectal Cancer and the American Cancer Society. *Gastroenterology* 2006, **130**(6):1872-1885.
2. Stryker SJ, Wolff BG, Culp CE, Libbe SD, Ilstrup DM, MacCarty RL: **Natural history of untreated colonic polyps**. *Gastroenterology* 1987, **93**(5):1009-1013.
3. Pinto D, Clevers H: **Wnt, stem cells and cancer in the intestine**. *Biol Cell* 2005, **97**(3):185-196.
4. Korinek V, Barker N, Morin PJ, van Wichen D, de Weger R, Kinzler KW, Vogelstein B, Clevers H: **Constitutive transcriptional activation by a beta-catenin-Tcf complex in APC-/- colon carcinoma**. *Science* 1997, **275**(5307):1784-1787.
5. Morin PJ, Sparks AB, Korinek V, Barker N, Clevers H, Vogelstein B, Kinzler KW: **Activation of beta-catenin-Tcf signaling in colon cancer by mutations in beta-catenin or APC**. *Science* 1997, **275**(5307):1787-1790.
6. Marra G, Jiricny J: **Multiple colorectal adenomas--is their number up?** *N Engl J Med* 2003, **348**(9):845-847.
7. **Update on the Paris classification of superficial neoplastic lesions in the digestive tract**. *Endoscopy* 2005, **37**(6):570-578.
8. Jass JR: **Classification of colorectal cancer based on correlation of clinical, morphological and molecular features**. *Histopathology* 2007, **50**(1):113-130.
9. di Pietro M, Bellver JS, Menigatti M, Bannwart F, Schnider A, Russell A, Truninger K, Jiricny J, Marra G: **Defective DNA mismatch repair determines a characteristic transcriptional profile in proximal colon cancers**. *Gastroenterology* 2005, **129**(3):1047-1059.
10. Ter Braak C: **Canonical Correspondence Analysis: A new Eigenvector technique for multivariate direct gradient analysis**. *Ecology* 1986, **67**(5):1167-1179.
11. Miki R, Kadota K, Bono H, Mizuno Y, Tomaru Y, Carninci P, Itoh M, Shibata K, Kawai J, Konno H *et al*: **Delineating developmental and metabolic pathways in vivo by expression profiling using the RIKEN set of 18,816 full-length enriched mouse cDNA arrays**. *Proc Natl Acad Sci U S A* 2001, **98**(5):2199-2204.
12. Wu LF, Hughes TR, Davierwala AP, Robinson MD, Stoughton R, Altschuler SJ: **Large-scale prediction of Saccharomyces cerevisiae gene function using overlapping transcriptional clusters**. *Nat Genet* 2002, **31**(3):255-265.
13. Zhang W, Morris QD, Chang R, Shai O, Bakowski MA, Mitsakakis N, Mohammad N, Robinson MD, Zirngibl R, Somogyi E *et al*: **The functional landscape of mouse gene expression**. *J Biol* 2004, **3**(5):21.
14. van de Wetering M, Barker N, Harkes IC, van der Heyden M, Dijk NJ, Hollestelle A, Klijn JG, Clevers H, Schutte M: **Mutant E-cadherin breast**

cancer cells do not display constitutive Wnt signaling. *Cancer Res* 2001, **61**(1):278-284.
15. van de Wetering M, Sancho E, Verweij C, de Lau W, Oving I, Hurlstone A, van der Horn K, Batlle E, Coudreuse D, Haramis AP *et al*: **The beta-catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells**. *Cell* 2002, **111**(2):241-250.
16. van de Wetering M, Oving I, Muncan V, Pon Fong MT, Brantjes H, van Leenen D, Holstege FC, Brummelkamp TR, Agami R, Clevers H: **Specific inhibition of gene expression using a stably integrated, inducible small-interfering-RNA vector**. *EMBO Rep* 2003, **4**(6):609-615.
17. Notterman DA, Alon U, Sierk AJ, Levine AJ: **Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays**. *Cancer Res* 2001, **61**(7):3124-3130.
18. Lin YM, Furukawa Y, Tsunoda T, Yue CT, Yang KC, Nakamura Y: **Molecular diagnosis of colorectal tumors by expression profiles of 50 genes expressed differentially in adenomas and carcinomas**. *Oncogene* 2002, **21**(26):4120-4128.
19. Nosho K, Yamamoto H, Adachi Y, Endo T, Hinoda Y, Imai K: **Gene expression profiling of colorectal adenomas and early invasive carcinomas by cDNA array analysis**. *Br J Cancer* 2005, **92**(7):1193-1200.
20. Nata K, Liu Y, Xu L, Ikeda T, Akiyama T, Noguchi N, Kawaguchi S, Yamauchi A, Takahashi I, Shervani NJ *et al*: **Molecular cloning, expression and chromosomal localization of a novel human REG family gene, REG III**. *Gene* 2004, **340**(1):161-170.
21. Bishnupuri KS, Luo Q, Murmu N, Houchen CW, Anant S, Dieckgraefe BK: **Reg IV activates the epidermal growth factor receptor/Akt/AP-1 signaling pathway in colon adenocarcinomas**. *Gastroenterology* 2006, **130**(1):137-149.
22. Hanai J, Mammoto T, Seth P, Mori K, Karumanchi SA, Barasch J, Sukhatme VP: **Lipocalin 2 diminishes invasiveness and metastasis of Ras-transformed cells**. *J Biol Chem* 2005, **280**(14):13641-13647.
23. Comijn J, Berx G, Vermassen P, Verschueren K, van Grunsven L, Bruyneel E, Mareel M, Huylebroeck D, van Roy F: **The two-handed E box binding zinc finger protein SIP1 downregulates E-cadherin and induces invasion**. *Mol Cell* 2001, **7**(6):1267-1278.
24. Jankowski JA, Bedford FK, Boulton RA, Cruickshank N, Hall C, Elder J, Allan R, Forbes A, Kim YS, Wright NA *et al*: **Alterations in classical cadherins associated with progression in ulcerative and Crohn's colitis**. *Lab Invest* 1998, **78**(9):1155-1167.
25. Hardy RG, Tselepis C, Hoyland J, Wallis Y, Pretlow TP, Talbot I, Sanders DS, Matthews G, Morton D, Jankowski JA: **Aberrant P-cadherin**

expression is an early event in hyperplastic and dysplastic transformation in the colon. *Gut* 2002, **50**(4):513-519.
26. Lin SY, Elledge SJ: **Multiple tumor suppressor pathways negatively regulate telomerase**. *Cell* 2003, **113**(7):881-889.
27. Mariadason JM, Nicholas C, L'Italien KE, Zhuang M, Smartt HJ, Heerdt BG, Yang W, Corner GA, Wilson AJ, Klampfer L *et al*: **Gene expression profiling of intestinal epithelial cell maturation along the crypt-villus axis**. *Gastroenterology* 2005, **128**(4):1081-1088.
28. Ding H, Schertzer M, Wu X, Gertsenstein M, Selig S, Kammori M, Pourvali R, Poon S, Vulto I, Chavez E *et al*: **Regulation of murine telomere length by Rtel: an essential gene encoding a helicase-like protein**. *Cell* 2004, **117**(7):873-886.
29. Pitti RM, Marsters SA, Lawrence DA, Roy M, Kischkel FC, Dowd P, Huang A, Donahue CJ, Sherwood SW, Baldwin DT *et al*: **Genomic amplification of a decoy receptor for Fas ligand in lung and colon cancer**. *Nature* 1998, **396**(6712):699-703.
30. Bai C, Connolly B, Metzker ML, Hilliard CA, Liu X, Sandig V, Soderman A, Galloway SM, Liu Q, Austin CP *et al*: **Overexpression of M68/DcR3 in human gastrointestinal tract tumors independent of gene amplification and its location in a four-gene cluster**. *Proc Natl Acad Sci U S A* 2000, **97**(3):1230-1235.
31. Postma C, Hermsen MA, Coffa J, Baak JP, Mueller JD, Mueller E, Bethke B, Schouten JP, Stolte M, Meijer GA: **Chromosomal instability in flat adenomas and carcinomas of the colon**. *J Pathol* 2005, **205**(4):514-521.
32. Thiebault K, Mazelin L, Pays L, Llambi F, Joly MO, Scoazec JY, Saurin JC, Romeo G, Mehlen P: **The netrin-1 receptors UNC5H are putative tumor suppressors controlling cell death commitment**. *Proc Natl Acad Sci U S A* 2003, **100**(7):4173-4178.
33. Dhanabal M, Jeffers M, LaRochelle WJ, Lichenstein HS: **Angioarrestin: a unique angiopoietin-related protein with anti-angiogenic properties**. *Biochem Biophys Res Commun* 2005, **333**(2):308-315.
34. Backlund MG, Mann JR, Holla VR, Buchanan FG, Tai HH, Musiek ES, Milne GL, Katkuri S, DuBois RN: **15-Hydroxyprostaglandin dehydrogenase is down-regulated in colorectal cancer**. *J Biol Chem* 2005, **280**(5):3217-3223.
35. Hogan MC, Griffin MD, Rossetti S, Torres VE, Ward CJ, Harris PC: **PKHDL1, a homolog of the autosomal recessive polycystic kidney disease gene, encodes a receptor with inducible T lymphocyte expression**. *Hum Mol Genet* 2003, **12**(6):685-698.
36. Nakayama M, Kikuno R, Ohara O: **Protein-protein interactions between large proteins: two-hybrid screening using a functionally classified library composed of long cDNAs**. *Genome Res* 2002, **12**(11):1773-1784.

37. Tamagnone L, Artigiani S, Chen H, He Z, Ming GI, Song H, Chedotal A, Winberg ML, Goodman CS, Poo M *et al*: **Plexins are a large family of receptors for transmembrane, secreted, and GPI-anchored semaphorins in vertebrates**. *Cell* 1999, **99**(1):71-80.
38. Comoglio PM, Tamagnone L, Giordano S: **Invasive growth: a two-way street for semaphorin signalling**. *Nat Cell Biol* 2004, **6**(12):1155-1157.
39. Michishita E, Garces G, Barrett JC, Horikawa I: **Upregulation of the KIAA1199 gene is associated with cellular mortality**. *Cancer Lett* 2006, **239**(1):71-77.
40. Horikawa I, Parker ES, Solomon GG, Barrett JC: **Upregulation of the gene encoding a cytoplasmic dynein intermediate chain in senescent human cells**. *J Cell Biochem* 2001, **82**(3):415-421.
41. Abe S, Katagiri T, Saito-Hisaminato A, Usami S, Inoue Y, Tsunoda T, Nakamura Y: **Identification of CRYM as a candidate responsible for nonsyndromic deafness, through cDNA microarray analysis of human cochlear and vestibular tissues**. *Am J Hum Genet* 2003, **72**(1):73-82.
42. Abe S, Usami S, Nakamura Y: **Mutations in the gene encoding KIAA1199 protein, an inner-ear protein expressed in Deiters' cells and the fibrocytes, as the cause of nonsyndromic hearing loss**. *J Hum Genet* 2003, **48**(11):564-570.
43. Truninger K, Menigatti M, Luz J, Russell A, Haider R, Gebbers JO, Bannwart F, Yurtsever H, Neuweiler J, Riehle HM *et al*: **Immunohistochemical analysis reveals high frequency of PMS2 defects in colorectal cancer**. *Gastroenterology* 2005, **128**(5):1160-1171.
44. di Pietro M, Marra G, Cejka P, Stojic L, Menigatti M, Cattaruzza MS, Jiricny J: **Mismatch repair-dependent transcriptome changes in human cells treated with the methylating agent N-methyl-n'-nitro-N-nitrosoguanidine**. *Cancer Res* 2003, **63**(23):8158-8166.
45. Benzécri J-P: **Correspondence Analysis Handbook**. New York: Marcel Dekker Inc. ; 1992.
46. Lee HK, Braynen W, Keshav K, Pavlidis P: **ErmineJ: tool for functional analysis of gene expression data sets**. *BMC Bioinformatics* 2005, **6**:269.
47. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT *et al*: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium**. *Nat Genet* 2000, **25**(1):25-29.
48. Barker N, Hurlstone A, Musisi H, Miles A, Bienz M, Clevers H: **The chromatin remodelling factor Brg-1 interacts with beta-catenin to promote target gene activation**. *Embo J* 2001, **20**(17):4935-4943.

49. Maake C, Auf der Maur F, Jovanovic K, Reinecke M, Hauri D, John H: **Occurrence and localization of uroguanylin in the aging human prostate**. *Histochem Cell Biol* 2003, **119**(1):69-76.
50. Kumar S, Tamura K, Nei M: **MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment**. *Brief Bioinform* 2004, **5**(2):150-163.

Figure Legends

Figure 1: Unsupervised analyses of microarray data

(**A**) *Hierarchical clustering analysis*: The 64 tissue samples represented on the *x*-axis include 32 normal mucosal samples (green branches) and 32 adenomas (red branches). Each probe plotted on the *y*-axis is color-coded to indicate the gene's level of expression relative to its median expression level across the entire tissue-sample set (blue: low; red: high). In the adenoma dendrogram, branches representing individual samples and small groups merge at higher levels than those of the normal-mucosa dendrogram, reflecting lower-level correlation, i.e., higher variability among the adenoma specimens. (**B**) *Principal component analysis (PCA).* Profile plot of the normalized first principal component (PCA-1) across the 64 specimens (green dots: normal mucosa; red dots: adenomas). The two tissue groups differ significantly in terms of PCA-1 ($p<0.0001$), which accounted for 26% of the total variance. Note the higher variability of the PCA-1 values in the adenoma group (higher fluctuation). (**C**) *Correlation analysis*. Tile-plot visualization of the pair-wise correlations of the samples. Correlation values are indicated on the gray-scale column (white > black: high > low). High correlation is observed among the samples

within each group (upper right quadrant: adenomas; lower left quadrant: normal mucosa) although the adenomas displayed somewhat greater diversity (i.e., on the whole, the gray tones in the upper right quadrant are darker than those in the lower left quadrant). Normal and adenoma samples are poorly correlated (upper left and lower right quadrants). However, samples from the same patient generally showed higher correlation than that observed between normal and adenoma samples from different patients (bright pixels on the secondary diagonals in the upper left and lower right quadrants). This finding probably reflects the strong influence of several factors, including the individual genetic background and life-style, and the fact that the normal and adenomatous tissues from a given patient were from the same colon segment. (**D**) *Correspondence analysis (CA)* of mRNA log(intensity) values of expressed genes from 27 of the 32 tissue pairs (normal mucosa: green dots, adenoma: red dots). The other five pairs were excluded from this analysis because one of the two samples behaved as an outlier. Limiting our analysis to the more homogeneous pairs facilitated the comparison of the gene expression profiles for the two tissue groups and allowed more reliable identification of clinical/pathological variables associated with profile scatter (see Figure 2). The areas delimited by the ellipses represent 95% of the estimated bi-normal distribution of the sample scores on the first and second CA axes. The map of the sample scores on the first two axes shows that CA efficiently discriminates between normal and adenoma samples. Higher variability is evident in the adenoma group, where the samples are more widely dispersed.

28

Figure 2: Clinical/pathological variables that correlate with distinct gene expression profiles. The panels summarize the most important results of the canonical correspondence analysis (CCA) of mRNA intensity log-ratio values (adenoma : normal) of expressed genes. For clarity, CCA axis 1 has been drawn vertically in both panels. A: Correlation between specific clinical/pathological variables (*adenoma diameter*, *colon segment of origin, degree of dysplasia, adenoma recurrence*) and clusters of differential gene expression profiles (coded as log-ratio profiles), like those shown in panel B. Each vector represents a specific value for a given variable (e.g., *adenoma diameter*: > 20 mm, *dysplasia*: high-degree, etc.) and points toward the center of the profile cluster correlated with the clinical/pathological characteristic it represents. If the centers for each specific value are separated, the corresponding vectors point in distinct directions; otherwise, they are directed toward the same point. In the former case, the represented variable can be assumed to be significantly correlated with the profiles; in the latter case, there is no correlation. The length of the vector reflects the strength of the correlation: those approaching the circumference of the correlation circle, which represents a correlation value of 1, indicate stronger correlation than shorter vectors (correlation closer to 0). (Abbreviations: d, diameter; Hd, high-degree dysplasia; Ld, low-degree dysplasia; A, ascending colon ; T, transverse colon; D, descending colon; S, sigmoid colon; R, rectum; Rec, recurrent adenomas; no Rec, no recurrent adenomas). Unlabeled vectors are related to variables that were not clearly associated with any distinct

cluster of expression profiles. Larger adenomas were predictably associated with high-degree dysplasia. In contrast, their association with non-recurrence was unexpected and probably due to the fact that patients who had already undergone endoscopic polypectomy (i.e., those with recurrence) presented relatively recent-onset (consequently, smaller) polyps at the study colonoscopy. **B.** CCA score plot with samples grouped by *adenoma diameter*. Each of the three size-related groups is delimited by an ellipse with the center labeled. The ellipse representing the adenomas measuring > 20 mm in diameter shows very little overlap with those of the other two groups (adenomas with diameters of 20 mm and those with diameters < 20 mm).

***Figure 3: KIAA1199* is a putative target of Wnt signaling.**

(**A**) Degree of correlation between the expression of *KIAA1199* mRNA and that of 19 known Wnt-signaling target genes identified with the procedure described in Methods, Results, and Additional data file 8. For each of the 20 genes, the graph shows the normalized intensity of expression level (plotted on the *y*-axis) in each of the 32 adenomas and corresponding samples of normal mucosa (*x*-axis). (**B**) Mean expression of each gene in normal mucosa (green dots) and adenomas (red dots). Bars: confidence interval. (**C**) Northern blot showing reduced *KIAA1199* expression in LS174T cells following doxycycline-mediated induction of β-catenin siRNA, dominant negative TCF4 (dnTCF4), or N-terminal-deleted TCF4 (N-TCF4). The ~ 8-Kb band corresponds to full-length *KIAA1199* mRNA. The lower band (~5 Kb) may

represent an alternative form of this mRNA. Dox: cell transfectants grown in the presence or absence of doxycycline; Tr1: a parental clone (i.e., cells expressing the repressor protein modified by doxycycline but not transfected with β-catenin siRNA, dnTCF4, or N-TCF4) used as a control of doxycycline exposure. As a loading control, the ethidium-bromide-stained agarose gel is shown in the lower panel.

Figure 4: Expression of KIAA1199 mRNA and protein in normal intestinal mucosa and colorectal tumors. *In situ* hybridization studies (panels A-C) localized *KIAA1199* mRNA expression to the lower portion of normal epithelial crypts (**A, B**) and revealed that expression is markedly upregulated in colorectal tumors (**C**). Note the different levels of expression in tumor glands and normal crypts (*). (**D**) KIAA1199 protein expression is also limited to the lower half of the normal colonic crypts, and a similar pattern is observed in the ileal mucosa (**E**), where the protein is expressed only in the crypts (not in the villi). In panels **F**) and **G**), adenomatous crypts with low-grade dysplasia present increased expression of KIAA1199, particularly in the cytoplasm facing the crypt lumen, and in and around the mucin vacuoles of goblet cells. (Note the striking difference with goblet cells of normal crypts in both panels.) The expression pattern changes dramatically during the transition from low-grade dysplasia with goblet-cell differentiation (**H**) to high-grade dysplasia in which this differentiation is no longer apparent (**I**, arrowheads: multilayer of unstained nuclei occupying more than the basal half of the dysplastic epithelium). In more advanced colon tumors (**J** and **K**), KIAA1199 overexpression is

maintained. Note that, in **J**) and **I**), the expression of KIAA1199 protein (like that of KIAA1199 mRNA – panel C) is highest in the luminal portion of the dysplastic glands. **L**) Normal mucosa, with the corresponding tumor in the inset. Negative control: KIAA1199 antibody pre-absorbed with the peptide used to immunize rabbits.

Figure 5. Phylogenetic tree of the proteins containing the TMEM2-homology domain found in the central region of KIAA1199. The tree was generated with MEGA3 [50] from the multiple sequence alignment shown in Additional data file 12. It was calculated with the Minimum Evolution algorithm and the JTT matrix. Positions with gaps were removed for calculation of pairwise distances. Node robustness was assessed using the bootstrap method with 100 resamplings. (Bootstrap values are shown at the nodes.) Two branches emerged, one comprising KIAA1199 and TMEM2, and the other with polyductin, fibrocystin L, and a number of other THD-containing proteins found in the ciliate *Tetrahymena thermophila*, which were apparently generated in a series of *Tetrahymena*-specific gene duplications. The N-terminal repeats of polyductin and fibrocystin L clustered together, as did the C-terminal repeats, suggesting that the intragenic duplication of the TH domain in the ancestor of polyductin and fibrocystin L occurred before the divergence of chordates and echinoderms (more details in Additional data file 11).

Table 1. Characteristics of the 28 patients with adenomatous polyps included in the study

Patient	Age (y)	Sex	Colon Segment involved	Max. Adenoma Diameter (mm)	Microscopic Appearance	Highest Degree of Dysplasia in the Adenoma ^	Degree of Dysplasia at Sampling Site ^	No. of Adenomas at Study Colonoscopy ¥	No. of Previously Excised Adenomas ‡	No. of Previous and/or Synchronous Hyperplastic Polyps	Familiarity for Colorectal Cancer (Relative, Onset Age)
GL *	49	M	D/S	10/10	T-V/T-V	H/L	H/L	9	4 ƒŽ	15 ˜	Mother, 70
PR	74	F	S	20	T-V	H	H	2	-	0	no
PC *	69	M	S/S	10/20	T/T-V	H/H	H/H	10	-	1	no
FP *	57	M	S/S	15/30	T/T-V	H/H	H/L	1	-	1	Mother, 69
CD *	71	M	T/R	15/10	T/T	H/L	H/L	2	7 ƒ	2	no
MA	65	M	R	15	T-V	L	L	2	-	0	no
ME	63	M	R	15	T-V	L	L	9	-	1	Father, 79
RA	64	F	A	15	T	L	L	1	7 ƒ	0	Sister, 68
PR	72	M	R	40	T-V	H	H	5	-	0	no
SD	56	M	A	15	T	H	H	1	1 ƒ	0	Mother, 83; Sister, 87
MP	38	M	S	15	T-V	L	L	2	-	0	Father, 79
MP	61	M	S	20	T	L	L	3	-	2	no
LG #	41	M	R	20	T-V	H	L	5 (2 serrated)	-	0	Father, 60
LS #	45	M	S	20	T	H	H	1	-	1	Father, 60
BG	58	M	D	15	T-V	L	L	2	-	1	no
PL	69	F	S	15	T-V	L	L	2	-	0	no
SMA	52	F	S	30	T	H	H	2	-	1	no
MR	58	F	D	20	T	L	L	2	-	0	no
GN	69	M	R	40	T	H	H	2	-	0	no
BA	69	M	S	30	T	L	L	6	-	0	no
PF	56	M	S	30	T	L	L	2	-	0	no
RC	55	F	A	30	T-V	L	L	12	3 ƒŽ	1	no
TMA	58	F	S	10	T	L	L	1	-	0	Mother, 85
NM	52	M	R	35	T-V	T1 §	H	1	-	0	no
MA	83	M	S	10	T-V	H	H	2	-	1	no
MM	50	M	S	30	T-V	H	H	2	-	0	Father's Brother, 65
NF	79	M	A	20	T-V	L	L	2	-	0	Mother, 70
PN	67	F	S	15	T-V	L	L	1	-	1	no

* two adenomas from these patients were analized.

LG and LS are brothers.

^ Low (L) vs. high-grade (H) dysplasia as defined by the World Health Organization Classification of Tumors of the Digestive System, Editorial and Consensu Conference in Lyon, France, November 6-9, 1999. International Agency for Research on Cancer (IARC); IARC Press, Lyon 2000.

§ Superficial submucosal invasion (T1). The tissue collected for microarray came from the adenomatous portion of the polyp.

¥ This number includes the adenoma(s) subjected to microarray analysis

‡ Total no. adenomas detected and excised during previous colonoscopies. (-): No previous colonoscopies

ƒ These cases were considered as recurrent adenomas for the canonical correspondence analysis (CCA). In two of them (Ž), the index colonoscopy was performed in a different center about 10 years before the study colonoscopy.

˜ Hyperplastic polyposis

M, male; F, female; A, ascending colon ; T, transversum; D, descending colon; S, sigmoid colon; R, rectum; T, tubular; T-V, tubulo-villous.

Table 2. Genes most likely to be involved in the development and evolution of colorectal adenomas (a subset of genes listed in Additional data file 9) subdivided by Gene Ontology category.

Gene Symbol	Gene Name	Fold Differences * ▲	▼
Regulation of Transcription			
NLF1	nuclear localized factor 1	33.1	
FOXQ1	forkhead box Q1	24.4	
MSX2	msh homeobox homolog 2	22.2	
ASCL2	achaete-scute complex-like 2	17.3	
MSX1	msh homeobox homolog 1	8.5	
IRX3	iroquois homeobox protein 3	8.4	
GRHL3	grainyhead-like 3	7.9	
TRIM29	tripartite motif-containing 29	7.4	
ETV4	ets variant gene 4 (E1A enhancer binding protein, E1AF)	5.4	
ARNTL2	aryl hydrocarbon receptor nuclear translocator-like 2	5.3	
TEAD4	TEA domain family member 4	5.2	
SP5	Sp5 transcription factor	5.2	
HES6	hairy and enhancer of split 6	4.6	
TBX3	T-box 3	4.6	
NFE2L3	nuclear factor (erythroid-derived 2)-like 3	4.3	
GRHL1	grainyhead-like 1	4.2	
FEV	FEV (ETS oncogene family)		15.1
SPIB	Spi-B transcription factor		13.2
NEUROD1	neurogenic differentiation 1		10.6
MEIS1	Meis1, myeloid ecotropic viral integration site 1		7.1
NR3C1	Nuclear receptor subfamily 3, group C, member 1		5.9
NR5A2	nuclear receptor subfamily 5, group A, member 2		5.6
THRB	thyroid hormone receptor, beta		5.2
ZNF483	Zinc finger protein 483		5.1
ZFHX1B	Zinc finger homeobox 1b (SIP-1)		4.8
MEOX2	mesenchyme homeobox 2		4.7
HOXD10	homeobox D10		4.6
MAF	v-maf musculoaponeurotic fibrosarcoma oncogene		4.5
SOX10	SRY (sex determining region Y)-box 10		4.2
Cell Proliferation / Differentiation / Apoptosis			
REG1B	regenerating islet-derived 1 beta	75.8	
REG3A	regenerating islet-derived 3 alpha	29.5	
TACSTD2	tumor-associated calcium signal transducer 2	21.4	
IL8	interleukin 8	14.7	
SERPINB5	serpin peptidase inhibitor, clade B, member 5 (Maspin)	13.8	
REG1A	regenerating islet-derived 1 alpha	8.2	
FAIM2	Fas apoptotic inhibitory molecule 2	7.5	
DUSP4	dual specificity phosphatase 4	7.4	
REG4	regenerating islet-derived family, member 4	6.8	
PHLDA1	pleckstrin homology-like domain, family A, member 1	6.0	
LCN2	lipocalin 2 (oncogene 24p3)	5.7	
RTEL1	regulator of telomere elongation helicase 1	5.6	
TGFBI	transforming growth factor, beta-induced	5.2	
IGFBP2	insulin-like growth factor binding protein 2	4.8	
TDGF1	teratocarcinoma-derived growth factor 1	4.7	
TNFRSF6B	tumor necrosis factor receptor superfamily, memb. 6b, decoy	4.5	
DMBT1	deleted in malignant brain tumors 1	4.2	
TNFRSF10C	tumor necrosis factor receptor superfamily, member 10c, decoy	4.1	
ANGPTL1	angiopoietin-like 1 (Angioarrestin)		24.9
CDKN2B	cyclin-dependent kinase inhibitor 2B (p15, inhibits CDK4)		14.9
GPM6B	glycoprotein M6B		11.5
ANK2	ankyrin 2		9.8
UNC5C	unc-5 homolog C		7.4
HPGD	Hydroxyprostaglandin dehydrogenase 15-(NAD)		6.1
CPNE8	copine VIII		5.5
FAIM3	Fas apoptotic inhibitory molecule 3		5.4
IL6R	interleukin 6 receptor		4.8
TUSC3	tumor suppressor candidate 3		4.7
DUSP1	dual specificity phosphatase 1		4.7
Table 2, page 2			
RERG	RAS-like, estrogen-regulated, growth inhibitor		4.6
NDN	necdin		4.5
IGF1	insulin-like growth factor 1 (somatomedin C)		4.0
Cell Adhesion			
CDH3	cadherin 3, type 1, P-cadherin	81.7	
CLDN2	claudin 2	16.1	
CLDN1	claudin 1	9.0	
DSG3	desmoglein 3	7.3	
DSC3	desmocollin 3	7.2	
DSG4	desmoglein 4	5.9	
CLDN8	claudin 8		25.8
CDH19	cadherin 19, type 2		8.3
CEACAM7	carcinoembryonic antigen-related cell adhesion molecule 7		8.3
CLDN23	claudin 23		8.0
NRXN1	neurexin 1		7.1
PCDH19	protocadherin 19		6.8
NLGN4X	neuroligin 4, X-linked		6.0
TNXB	tenascin XB		5.6
MUCDHL	mucin and cadherin-like		5.1
PCDH9	protocadherin 9		4.9
L1CAM	L1 cell adhesion molecule		4.2

* Overexpresed (▲) or underexpressed (▼) in adenomas (vs. normal mucosa samples)

Additional data files

Additional data file 1 contains Supplementary Figure 1, which shows the unsupervised hierarchical clustering of the normal colonic mucosa, colorectal adenomas, and colorectal carcinomas. Additional data file 2 contains Supplementary Table 1 with the gene-ontology biological-process categories significantly over-represented in the set of genes differentially expressed in adenomas. Additional data file 3 contains Supplementary Figure 2 with tile plots showing the differential expression in adenomas (vs. normal mucosa) of genes involved in different biological processes. Additional data file 4 contains Supplementary Table 2 reporting the differential expression of known Wnt signaling genes in normal colorectal mucosa and colorectal adenomas. Additional data file 5 contains Supplementary Table 3, which reports a selection of genes whose expression displayed significant positive or negative correlation with adenoma size. Additional data file 6 contains Supplementary Figure 3 with data regarding the amplification of *KIAA1199* mRNA by real-time RT-PCR. Additional data file 7 contains Supplementary Table 4 with the list of known Wnt target genes. Additional data file 8 contains Supplementary Figure 4 showing the procedure used to analyze the degree of co-expression between *KIAA1199* and known Wnt target genes. Additional data file 9 contains Supplementary Table 5, which reports the list of genes with markedly altered expression in colorectal adenomas (fold differences of 4 or more compared with normal mucosa). Additional data file 10 contains Supplementary Table 6 reporting the transcript level in adenomas of genes classified as transcription regulators.

Additional data file 11 contains supplementary text presenting a detailed *in silico* analysis of the KIAA1199 protein. Additional data file 12 contains Supplementary Figure 5 showing the sequence alignment of the TH domain for the region 406-1087 of the human KIAA1199 protein.

36

Figure 1. Sabates-Bellver et al.

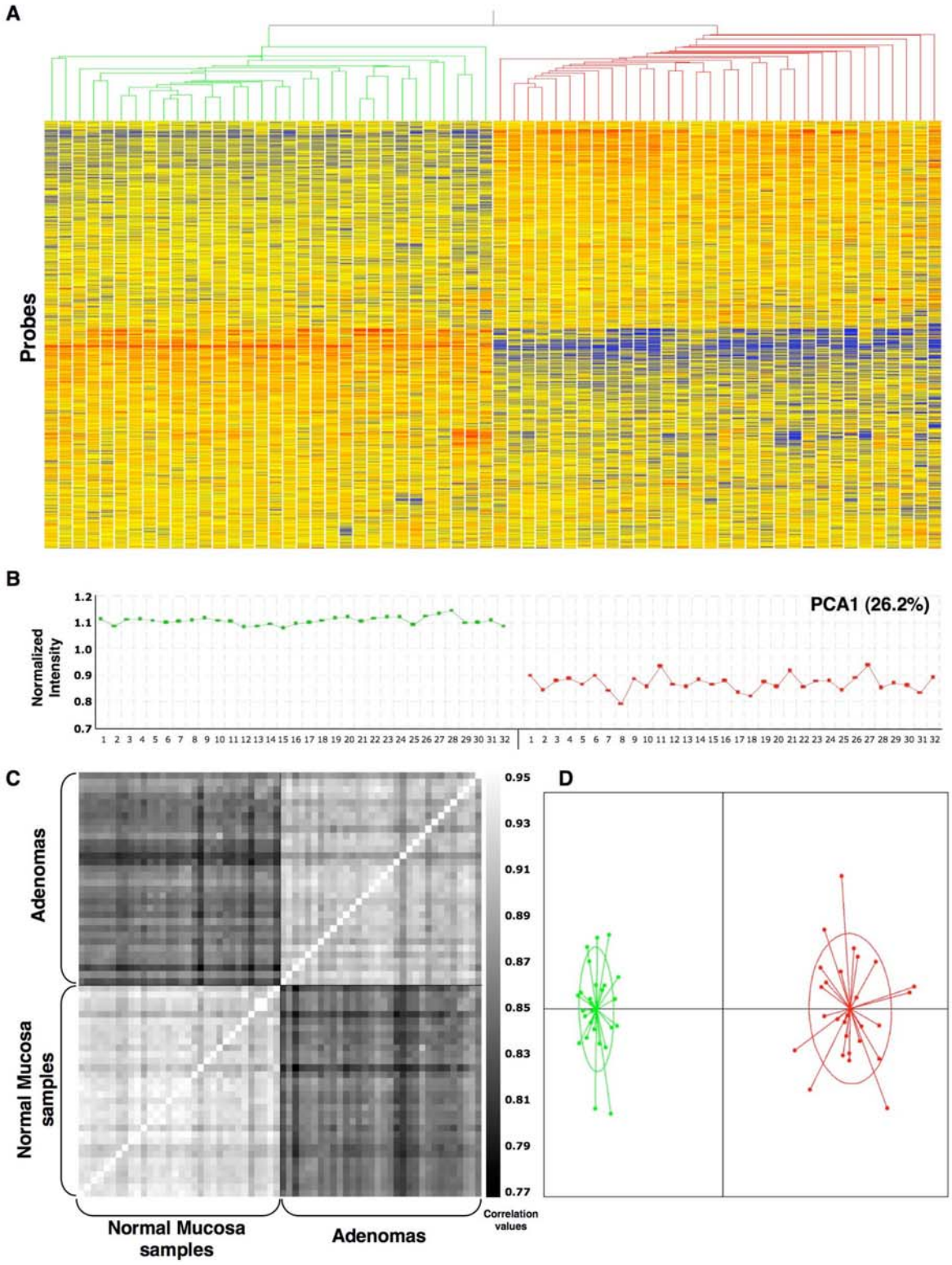


Figure 2. Sabates-Bellver et al.

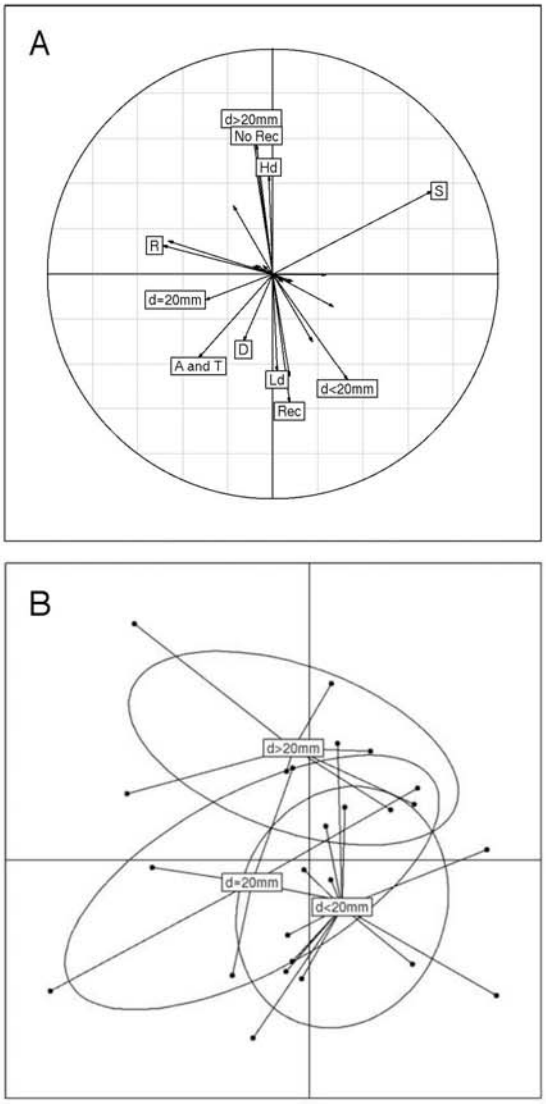


Figure 3, Sabates-Bellver et al.

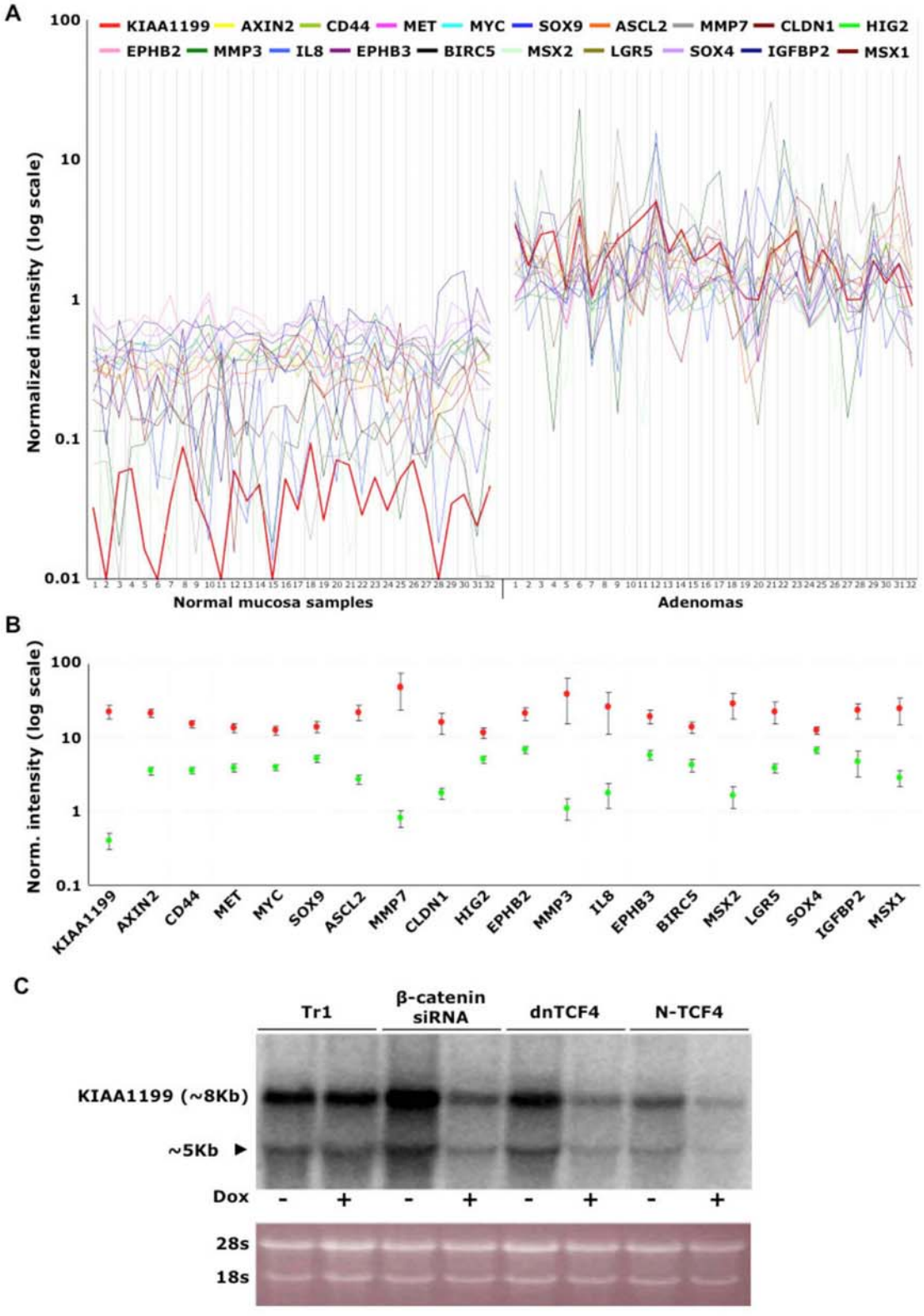


Figure 3

Figure 4, Sabates-Bellver et al.

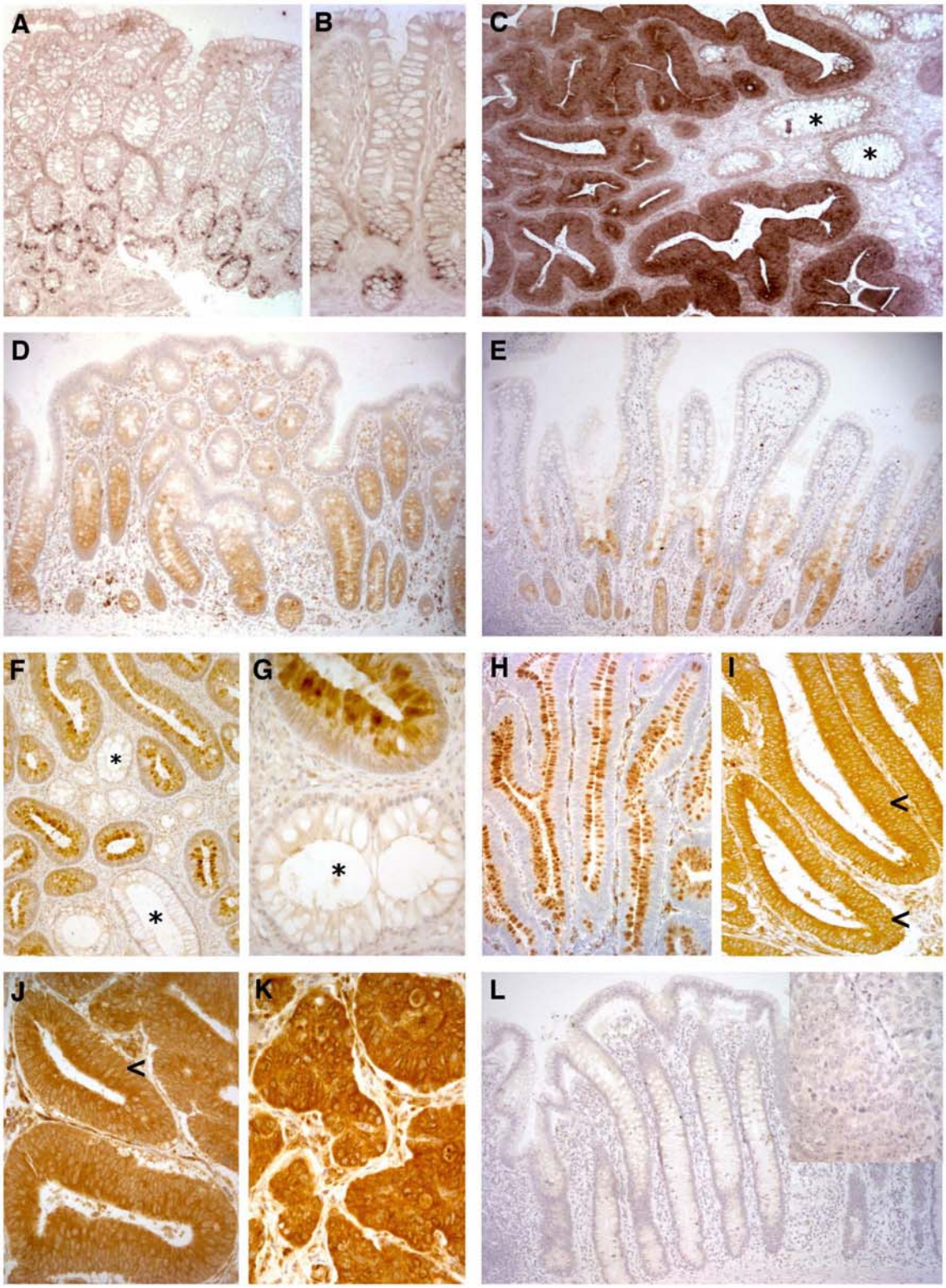


Figure 4

Figure 5, Sabates-Bellver et al.

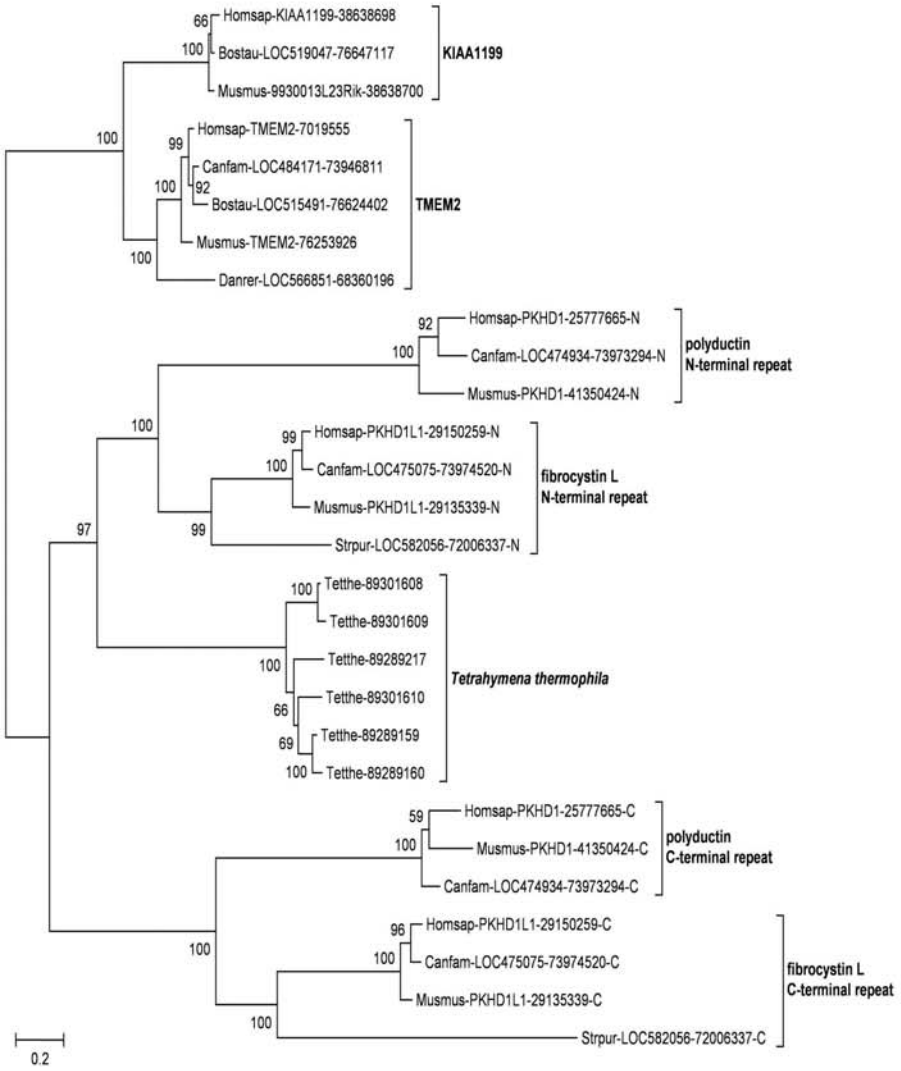
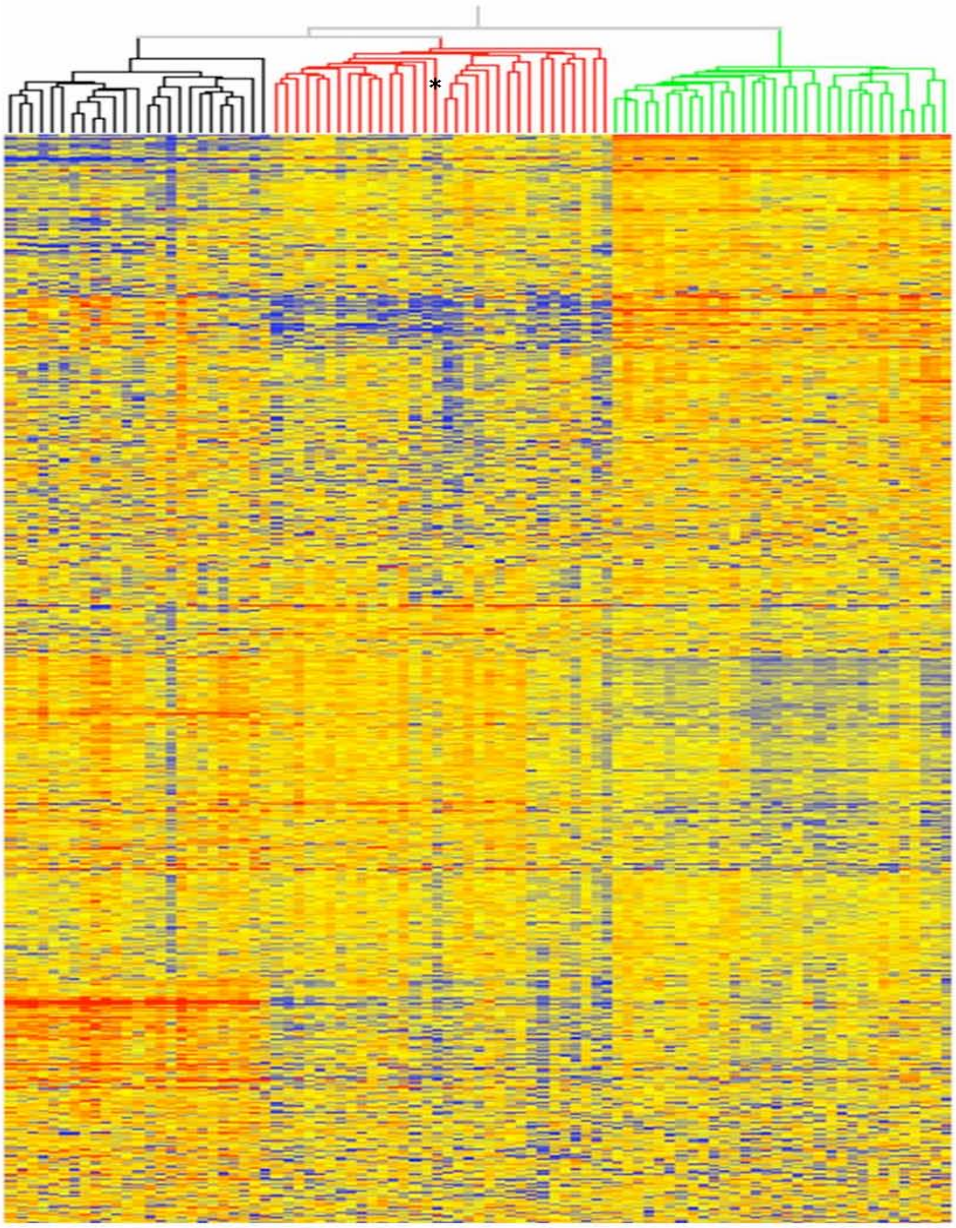


Figure 5

Supplementary Figure 1. Sabates-Bellver et al.



Supplementary Figure 1. **Unsupervised hierarchical clustering of normal colonic mucosa (32 samples, green branches), colorectal adenomas (32 samples, red branches), and advanced colorectal carcinomas (25 samples, black branches) based on transcriptome analysis with Affymetrix U133 Plus 2.0 arrays.** Data were analyzed as described in Methods. The 32,813 probes plotted on the *y*-axis (i.e., those expressed in at least one of the three tissue groups) are color-coded to indicate their expression levels relative to the median level for the gene across the entire sample set (blue: lower; red: higher). As expected, the tissue sample collected from patient NM's polyp (*) clustered with the adenomas. This sample was clearly adenomatous although the polyp itself was classified as a T1 tumor (Table 1).

Supplementary Table 1: Biological-process categories most significantly over-represented in the set of genes differentially expressed in adenomas (vs. normal mucosa) with fold differences of 2.0 or more (ErmineJ's overrepresentation analysis)

Genes overexpressed in adenomas

Gene Ontology category	Category genes *	Altered category genes ^	p-value #
Mitosis	137	37	2.51E-20
Cell Division	168	38	2.54E-17
DNA Replication	130	28	3.16E-12
Spindle Organization and Biogenesis	17	11	1.07E-11
Phosphoinisitide-mediated signaling	29	11	4.18E-08
Nucleotide Biosynthesis	50	13	3.49E-07
Glutamine Family Amino Acid Metabolism	14	7	7.97E-07
DNA Integrity Checkpoint	11	6	2.31E-06
Microtubule Cytoskeleton Organization and Biogenesis	29	9	4.30E-06
Purine Nucleotide Metabolism	13	6	8.54E-06
One-carbon compound metabolism	27	8	1.87E-05
Cytoplasm Organization and Biogenesis	10	5	2.20E-05
Inflammatory Response	151	20	4.72E-05
Regulation of Apoptosis	114	16	5.15E-05

Genes underexpressed in adenomas

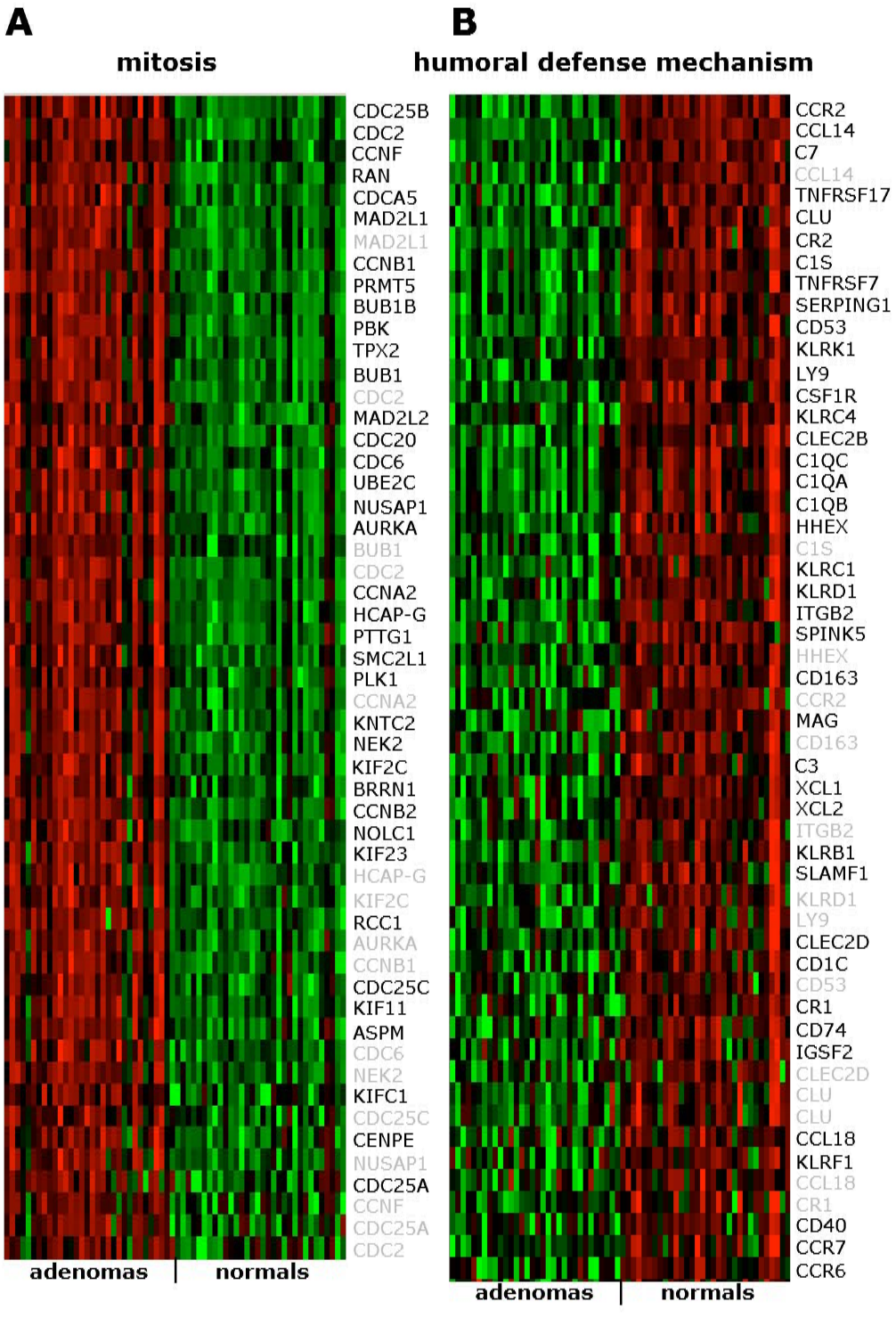
Gene Ontology category	Category genes *	Altered category genes ^	p-value #
Humoral Defense Mechanism	94	40	8.49E-19
Antimicrobial Humoral Response	72	31	1.47E-14
Inorganic Anion Transport	96	33	3.72E-12
Organ Development	190	45	6.94E-10
Inflammatory Response	151	35	3.35E-08
Circulation	77	24	3.70E-08
Elevation of Cytosolic Calcium Ion Concentration	29	14	4.46E-08
Enzyme Linked Receptor Protein Signaling Pathway	81	24	9.05E-08
Muscle Contraction	81	24	9.46E-08
Taxis	89	25	1.48E-07
Cell Motility	132	31	3.65E-07
Transmembrane Receptor Protein Tyrosine Kinase Signaling Pathway	109	26	2.67E-06
Cell-cell Signaling	186	36	5.54E-06
Cell-cell Adhesion	115	25	2.18E-05

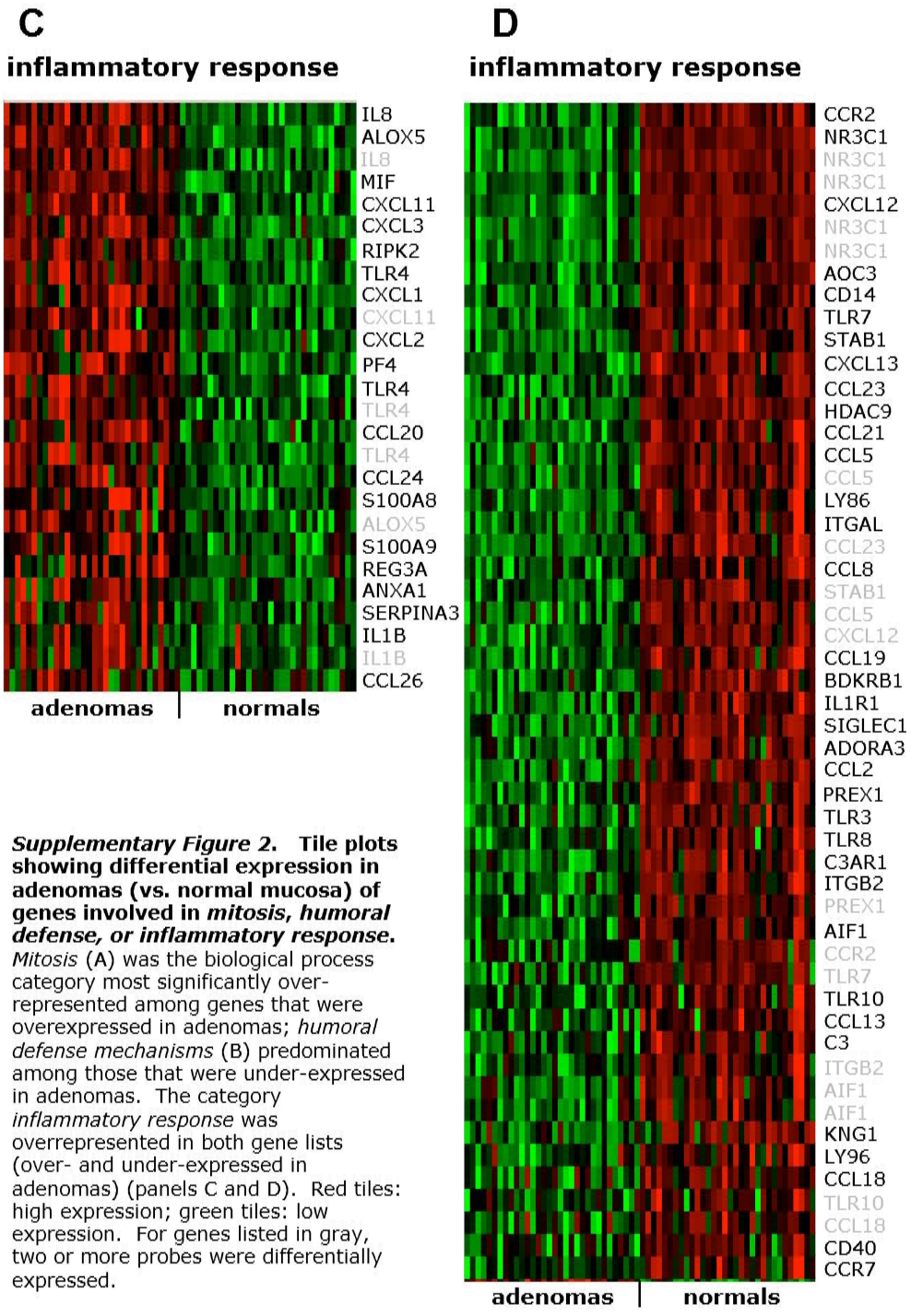
* Number of genes listed in the corresponding Gene Ontology category. Only categories containing 5-200 genes were considered.

^ Number of category genes that were significantly up- or down-regulated in adenomas

p-value corrected for multiple testing as reported in: Lee HK et al. "ErmineJ: Tool for functional analysis of gene expression data sets" *BMC Bioinformatics 2005; 6:269*

Supplementary Figure 2; Sabates-Bellver et al.





Supplementary Figure 2. **Tile plots showing differential expression in adenomas (vs. normal mucosa) of genes involved in *mitosis, humoral defense, or inflammatory response*.** *Mitosis* (A) was the biological process category most significantly over-represented among genes that were overexpressed in adenomas; *humoral defense mechanisms* (B) predominated among those that were under-expressed in adenomas. The category *inflammatory response* was overrepresented in both gene lists (over- and under-expressed in adenomas) (panels C and D). Red tiles: high expression; green tiles: low expression. For genes listed in gray, two or more probes were differentially expressed.

Supplementary Table 2. Expression of Wnt signaling genes# in normal colorectal mucosa and colorectal adenomas.

Gene Symbol §	Gene Name	Fold Differences Up*	Down*
Not expressed in normal mucosa and significantly changed in adenomas			
CST1	cystatin SN	6.9	
ETV4	ets variant gene 4 (E1A enhancer binding protein, E1AF)	5.4	
FGF18	Fibroblast growth factor 18	1.7	
FOSL1	FOS-like antigen 1	2.6	
IRX3	iroquois homeobox protein 3	8.4	
LOC283859	hypothetical protein LOC283859	10.5	
MMP3	matrix metallopeptidase 3 (stromelysin 1, progelatinase)	33.9	
MMP7	matrix metallopeptidase 7 (matrilysin, uterine)	56.9	
MSX1	msh homeobox homolog 1 (Drosophila)	8.5	
MSX2	msh homeobox homolog 2 (Drosophila)	22.2	
SMAD9	SMAD, mothers against DPP homolog 9 (Drosophila)	2.7	
SP5	Sp5 transcription factor	5.2	
TNFSF9	tumor necrosis factor (ligand) superfamily, member 9	2.4	
WNT2	wingless-type MMTV integration site family member 2	1.6	
Expressed in normal mucosa and changed in adenomas			
APCDD1	adenomatosis polyposis coli down-regulated 1	7.2	
ASCL2	achaete-scute complex-like 2 (Drosophila)	17.3	
AXIN2	axin 2 (conductin, axil)	7.3	
BIRC5	baculoviral IAP repeat-containing 5 (survivin)	3.2	
BMP4	bone morphogenetic protein 4	2.0	
CCND1	cyclin D1	2.0	
CCND2	cyclin D2	2.6	
CD44	CD44 molecule	4.4	
CLDN1	claudin 1	9.0	
CSNK1E	Casein kinase 1, epsilon	2.4	
CSNK2A2	casein kinase 2, alpha prime polypeptide	1.5	
CSNK2B	casein kinase 2, beta polypeptide	1.2	
CXXC4	CXXC finger 4	1.7	
EDAR	ectodysplasin A receptor	1.7	
EPHA1	EPH receptor A1	1.9	
EPHA2	EPH receptor A2	1.5	
EPHB2	EPH receptor B2	3.1	
EPHB3	EPH receptor B3	3.3	
EPHB4	EPH receptor B4	2.2	
FLJ43934	hypothetical protein LOC644544	1.4	
FUSIP1	FUS interacting protein (serine/arginine-rich) 1	1.4	
FZD3	frizzled homolog 3 (Drosophila)	2.8	
FZD6	frizzled homolog 6 (Drosophila)	1.8	
FZD7	frizzled homolog 7 (Drosophila)	1.6	
GSK3B	glycogen synthase kinase 3 beta	1.4	
HDGF	hepatoma-derived growth factor (high-mobility group protein 1-like)	1.6	

Supplementary Table 2 (page 2)

HIG2	hypoxia-inducible protein 2	2.3	
HMGN3	high mobility group nucleosomal binding domain 3	1.3	
IGFBP2	insulin-like growth factor binding protein 2, 36kDa	4.8	
IL8	interleukin 8	14.7	
LAMC2	laminin, gamma 2	2.0	
LEF1	lymphoid enhancer-binding factor 1	1.5	
LGR5	leucine-rich repeat-containing G protein-coupled receptor 5	5.8	
LOC646927	Hypothetical protein LOC646927	1.5	
MET	met proto-oncogene (hepatocyte growth factor receptor)	3.4	
MSLN	mesothelin	4.7	
MYC	v-myc myelocytomatosis viral oncogene homolog (avian)	3.2	
NKD2	naked cuticle homolog 2 (Drosophila)	1.7	
ODC1	ornithine decarboxylase 1	2.3	
PITX2	paired-like homeodomain transcription factor 2	1.5	
POP1	processing of precursor 1, ribonuclease P/MRP subunit	1.4	
PTTG1	pituitary tumor-transforming 1	2.3	
PTTG3	pituitary tumor-transforming 3	1.6	
RARG	retinoic acid receptor, gamma	1.8	
SMAD6	SMAD, mothers against DPP homolog 6 (Drosophila)	1.7	
SOX4	SRY (sex determining region Y)-box 4	2.8	
SOX9	SRY (sex determining region Y)-box 9	2.7	
TCF3	transcription factor 3 (E2A immunoglob. enhancer bind. fact. E12/E47)	2.3	
TCF7	transcription factor 7 (T-cell specific, HMG-box)	2.0	
TLE1	transducin-like enhancer of split 1 (E(sp1) homolog, Drosophila)	1.6	
WNT5A	wingless-type MMTV integration site family, member 5A	1.8	
ABCB1	ATP-binding cassette, sub-family B (MDR/TAP), member 1		2.6
APC	Adenomatosis polyposis coli		2.2
APC2	Adenomatosis polyposis coli 2		1.3
C2orf31	chromosome 2 open reading frame 31		1.5
CCND3	cyclin D3		1.2
CDX1	caudal type homeobox transcription factor 1		1.2
CSPG2	chondroitin sulfate proteoglycan 2 (versican)		2.1
DAAM1	Dishevelled associated activator of morphogenesis 1		1.9
DAAM2	dishevelled associated activator of morphogenesis 2		2.9
DHRS9	dehydrogenase/reductase (SDR family) member 9		3.2
DKK3	dickkopf homolog 3 (Xenopus laevis)		1.5
DRCTNNB1A	down-regulated by Ctnnb1, a		2.5
EDA	ectodysplasin A		1.8
EDN3	endothelin 3		11.1
EDNRB	endothelin receptor type B		1.8
EFNA1	ephrin-A1		2.3
EFNA5	ephrin-A5		1.7
EGFR	Epidermal growth factor receptor (v-erb-b oncogene homolog, avian)		2.0
EMP1	epithelial membrane protein 1		2.3
ENPP2	ectonucleotide pyrophosphatase/phosphodiesterase 2 (autotaxin)		2.0
EPHA10	EPH receptor A10		1.5
EPHA3	EPH receptor A3		9.5
EPHA7	EPH receptor A7		5.0

Supplementary Table 2 (page 3)

EPHA8	EPH receptor A8	1.3
EPHB6	EPH receptor B6	1.4
FGF9	fibroblast growth factor 9 (glia-activating factor)	3.8
FN1	fibronectin 1	4.1
FOSL2	FOS-like antigen 2	1.5
FRZB	frizzled-related protein (*SFRP3*)	2.6
FZD1	frizzled homolog 1 (Drosophila)	2.0
FZD4	frizzled homolog 4 (Drosophila)	1.5
FZD5	frizzled homolog 5 (Drosophila)	2.5
FZD8	frizzled homolog 8 (Drosophila)	2.4
GCG	glucagon	68.5
HDGFRP3	Hepatoma-derived growth factor, related protein 3	1.7
HHEX	homeobox, hematopoietically expressed	2.2
IGF1	insulin-like growth factor 1 (somatomedin C)	4.0
ISLR	immunoglobulin superfamily containing leucine-rich repeat	2.4
KLF5	Kruppel-like factor 5 (intestinal)	1.5
L1CAM	L1 cell adhesion molecule	4.2
LRRFIP2	Leucine rich repeat (in FLII) interacting protein 2	2.6
MAPK10	mitogen-activated protein kinase 10	1.9
MGC39518	hypothetical protein MGC39518	1.9
MITF	microphthalmia-associated transcription factor	2.2
NIN	ninein (GSK3B interacting protein)	1.5
NLK	Nemo-like kinase	1.6
POSTN	periostin, osteoblast specific factor	3.8
PPARD	peroxisome proliferative activated receptor, delta	1.8
PPP2CB	protein phosphatase 2 (formerly 2A), catalytic subunit, beta isoform	1.3
PRKCB1	Protein kinase C, beta 1	4.1
PRKCD	protein kinase C, delta	1.8
PRKCE	Protein kinase C, epsilon	1.5
PRKCH	Protein kinase C, eta	1.7
PRKCZ	protein kinase C, zeta	1.2
PTTG1IP	pituitary tumor-transforming 1 interacting protein	1.6
RET	ret proto-oncogene	1.6
RHOU	ras homolog gene family, member U	2.2
SFRP1	secreted frizzled-related protein 1	7.6
SFRP2	secreted frizzled-related protein 2	3.9
SLC9A3R1	solute carrier family 9, member 3 regulator 1	2.7
SMAD1	SMAD, mothers against DPP homolog 1 (Drosophila)	1.7
SMAD3	SMAD, mothers against DPP homolog 3 (Drosophila)	1.5
SMAD4	SMAD, mothers against DPP homolog 4 (Drosophila)	1.5
SNAI2	snail homolog 2 (Drosophila)	3.8
SNF1LK2	SNF1-like kinase 2	1.6
SOSTDC1	sclerostin domain containing 1	2.9
TCF2	transcription factor 2, hepatic; LF-B3; variant hepatic nuclear factor	2.1
TCF4	transcription factor 4	2.5
TCF7L1	transcription factor 7-like 1 (T-cell specific, HMG-box)	1.8
TCF7L2	Transcription factor 7-like 2 (T-cell specific, HMG-box)	1.7
TIAM1	T-cell lymphoma invasion and metastasis 1	2.4

Supplementary Table 2 (page 4)

TLE3	transducin-like enhancer of split 3 (E(sp1) homolog, Drosophila)	1.6
TLE4	transducin-like enhancer of split 4 (E(sp1) homolog, Drosophila)	1.6
VEGFC	vascular endothelial growth factor C	1.5
WNT2B	wingless-type MMTV integration site family, member 2B	2.6
WNT5B	wingless-type MMTV integration site family, member 5B	1.9
WNT9A	Wingless-type MMTV integration site family, member 9A	1.6
ZNF467	zinc finger protein 467	1.7
ZNF521	zinc finger protein 521	1.7

Expressed in normal mucosa and unchanged in adenomas

AES	amino-terminal enhancer of split
ATOH1	atonal homolog 1 (Drosophila)
AXIN1	axin 1
AXUD1	AXIN1 up-regulated 1
BCL9	B-cell CLL/lymphoma 9
BGLAP	bone gamma-carboxyglutamate (gla) protein (osteocalcin)
BTRC	beta-transducin repeat containing
C9orf126	Chromosome 9 open reading frame 126
CD24	CD24 molecule
CDH1	cadherin 1, type 1, E-cadherin (epithelial)
CSNK1A1	Casein kinase 1, alpha 1
CSNK1D	casein kinase 1, delta
CSNK1G1	casein kinase 1, gamma 1
CSNK1G2	casein kinase 1, gamma 2
CSNK1G3	casein kinase 1, gamma 3
CSNK2A1	casein kinase 2, alpha 1 polypeptide
CTBP1	C-terminal binding protein 1
CTBP2	C-terminal binding protein 2
CTGF	connective tissue growth factor
CTNNB1	catenin (cadherin-associated protein), beta 1, 88kDa
CTNNBIP1	catenin, beta interacting protein 1
CYR61	cysteine-rich, angiogenic inducer, 61
DIXDC1	DIX domain containing 1
DLL1	delta-like 1 (Drosophila)
DLL4	delta-like 4 (Drosophila)
DVL1	dishevelled, dsh homolog 1 (Drosophila)
DVL2	dishevelled, dsh homolog 2 (Drosophila)
DVL3	dishevelled, dsh homolog 3 (Drosophila)
EDN1	endothelin 1
EDNRA	endothelin receptor type A
EFNA3	ephrin-A3
EFNA4	ephrin-A4
EFNB1	ephrin-B1
EFNB2	ephrin-B2
EP300	E1A binding protein p300
EPHA4	EPH receptor A4
EPHA5	EPH receptor A5
EPHB1	EPH receptor B1
FBXW11	F-box and WD-40 domain protein 11

Supplementary Table 2 (page 5)

FBXW2	F-box and WD-40 domain protein 2
FBXW4	F-box and WD-40 domain protein 4
FGF4	fibroblast growth factor 4 (Kaposi sarcoma oncogene)
FOS	v-fos FBJ murine osteosarcoma viral oncogene homolog
FRAT1	frequently rearranged in advanced T-cell lymphomas
FRAT2	frequently rearranged in advanced T-cell lymphomas 2
FZD2	frizzled homolog 2 (Drosophila)
GJA1	gap junction protein, alpha 1, 43kDa (connexin 43)
GPR135	G protein-coupled receptor 135
GSK3A	glycogen synthase kinase 3 alpha
HSPG2	heparan sulfate proteoglycan 2 (perlecan)
ID2	inhibitor of DNA binding 2, dominant negative helix-loop-helix protein
ING1	inhibitor of growth family, member 1
JUN	v-jun sarcoma virus 17 oncogene homolog (avian)
JUP	junction plakoglobin
KREMEN1	kringle containing transmembrane protein 1
KRTHB6	keratin, hair, basic, 6 (monilethrix)
LDLR	low density lipoprotein receptor (familial hypercholesterolemia)
LOC642558	similar to FUS interacting protein (serine-arginine rich) 1
LRP5	low density lipoprotein receptor-related protein 5
LRP6	low density lipoprotein receptor-related protein 6
MAPK9	mitogen-activated protein kinase 9
MAPRE1	microtubule-associated protein, RP/EB family, member 1
MARK4	MAP/microtubule affinity-regulating kinase 4
MYCBP	c-myc binding protein
NFE2L2	nuclear factor (erythroid-derived 2)-like 2
OVOL1	ovo-like 1(Drosophila)
PAFAH1B1	platelet-activating factor acetylhydrolase, isoform Ib, alpha sub.45kDa
PLA2G2A	phospholipase A2, group IIA (platelets, synovial fluid)
PLAU	plasminogen activator, urokinase
PLAUR	plasminogen activator, urokinase receptor
PORCN	porcupine homolog (Drosophila)
PPARG	peroxisome proliferative activated receptor, gamma
PPP2CA	protein phosphatase 2, catalytic subunit, alpha isoform
PPP2R1A	protein phosphatase 2, regulatory subunit A (PR 65), alpha isoform
PPP2R1B	protein phosphatase 2, regulatory subunit A (PR 65), beta isoform
PPP2R5C	protein phosphatase 2, regulatory subunit B (B56), gamma isoform
PPP2R5E	protein phosphatase 2, regulatory subunit B (B56), epsilon isoform
PRKCA	protein kinase C, alpha
PRKCG	protein kinase C, gamma
PRKCI	protein kinase C, iota
PRKCQ	protein kinase C, theta
PRKD1	protein kinase D1
PTGS2	prostaglandin-endoperoxide synthase 2
PYGO2	pygopus homolog 2 (Drosophila)
RAC1	ras-related C3 botulinum toxin substrate 1 (rho family)
RHOV	ras homolog gene family, member V
RNF4	ring finger protein 4

Supplementary Table 2 (page 6)

SEC23B	Sec23 homolog B (S. cerevisiae)
SEMA3C	semaphorin 3C
SENP2	SUMO1/sentrin/SMT3 specific peptidase 2
SMAD2	SMAD, mothers against DPP homolog 2 (Drosophila)
SMAD5	SMAD, mothers against DPP homolog 5 (Drosophila)
SMAD7	SMAD, mothers against DPP homolog 7 (Drosophila)
SMO	smoothened homolog (Drosophila)
STRA6	stimulated by retinoic acid gene 6 homolog (mouse)
T	T, brachyury homolog (mouse)
TCF1	transcription factor 1, hepatic; LF-B1, hepatic nuclear factor (HNF1)
TLE2	transducin-like enhancer of split 2 (E(sp1) homolog, Drosophila)
TNFSF11	tumor necrosis factor (ligand) superfamily, member 11
TWIST1	twist homolog 1 (acrocephalosyndactyly 3; Saethre-Chotzen synd.)
UBXD1	UBX domain containing 1
UBXD2	UBX domain containing 2
UBXD3	UBX domain containing 3
UBXD4	UBX domain containing 4
USP33	ubiquitin specific peptidase 33
VEGF	vascular endothelial growth factor
VEGFB	vascular endothelial growth factor B
WNT6	wingless-type MMTV integration site family, member 6
ZNF651	zinc finger protein 651

Not expressed in both normal mucosa and adenomas

CDX4	caudal type homeobox transcription factor 4
CSNK1A1	Casein kinase 1, alpha 1
CX40.1	connexin40.1
DKK1	dickkopf homolog 1 (Xenopus laevis)
DKK2	dickkopf homolog 2 (Xenopus laevis)
DKK4	dickkopf homolog 4 (Xenopus laevis)
DKKL1	dickkopf-like 1 (soggy)
DLK1	delta-like 1 homolog (Drosophila)
DLL3	delta-like 3 (Drosophila)
EDN2	endothelin 2
EFNA2	ephrin-A2
EFNB3	ephrin-B3
ELAC1	ElaC homolog 1 (E. coli)
EN1	engrailed homolog 1
EN2	engrailed homolog 2
EPHA6	EPH receptor A6
FGF20	fibroblast growth factor 20
FLJ14712	hypothetical protein FLJ14712
FLJ90166	hypothetical protein FLJ90166
FOSB	FBJ murine osteosarcoma viral oncogene homolog B
FOXN1	forkhead box N1
FSHB	follicle stimulating hormone, beta polypeptide
FST	follistatin
FZD10	frizzled homolog 10 (Drosophila)
FZD9	frizzled homolog 9 (Drosophila)

Supplementary Table 2 (page 7)

GAST	gastrin
GJA10	gap junction protein, alpha 10, 59kDa
GJB6	gap junction protein, beta 6 (connexin 30)
IGF2	insulin-like growth factor 2 (somatomedin A)
IL6	interleukin 6 (interferon, beta 2)
KREMEN2	kringle containing transmembrane protein 2
KRTAP2-1	keratin associated protein 2-1
LOC391241	similar to Soggy-1 protein precursor (SGY-1)
LOC541472	hypothetical LOC541472
LOC641518	hypothetical protein LOC641518
LOC644985	Hypothetical protein LOC644985
MMP26	matrix metallopeptidase 26
NEUROG1	neurogenin 1
NEUROG3	neurogenin 3
NKD1	naked cuticle homolog 1 (Drosophila)
NRCAM	neuronal cell adhesion molecule
PTTG2	pituitary tumor-transforming 2
RUNX2	runt-related transcription factor 2
SFRP4	secreted frizzled-related protein 4
SFRP5	secreted frizzled-related protein 5
SHFM3P1	split hand/foot malformation (ectrodactyly) type 3 pseudogene 1
SIX3	sine oculis homeobox homolog 3 (Drosophila)
SOX17	SRY (sex determining region Y)-box 17
SOX2	SRY (sex determining region Y)-box 2
TLE6	transducin-like enhancer of split 6 (E(sp1) homolog, Drosophila)
WIF1	WNT inhibitory factor 1
WISP1	WNT1 inducible signaling pathway protein 1
WISP2	WNT1 inducible signaling pathway protein 2
WISP3	WNT1 inducible signaling pathway protein 3
WNT1	wingless-type MMTV integration site family, member 1
WNT10A	wingless-type MMTV integration site family, member 10A
WNT10B	wingless-type MMTV integration site family, member 10B
WNT11	wingless-type MMTV integration site family, member 11
WNT16	wingless-type MMTV integration site family, member 16
WNT3	wingless-type MMTV integration site family, member 3
WNT4	wingless-type MMTV integration site family, member 4
WNT7A	wingless-type MMTV integration site family, member 7A
WNT7B	wingless-type MMTV integration site family, member 7B
WNT8A	wingless-type MMTV integration site family, member 8A
WNT8B	wingless-type MMTV integration site family, member 8B
WNT9B	wingless-type MMTV integration site family, member 9B

Those classified as Wnt-signaling genes by R. Nusse (http://www.stanford.edu/~rnusse/pathways/targets.html and http://www.stanford.edu/~rnusse/pathways/targetcomp.html) and/or in the Gene Ontology database. Genes in red are the Wnt-signaling targets reported in Supplementary Table 4.

§ Within each category, genes are listed alphabetically by symbol
* Significantly overexpressed (Up) or underexpressed (Down) in adenomas (vs. normal mucosa)

Supplementary Table 3: A selection of genes whose expression displayed significant positive or negative correlation with adenoma size, i.e., increase or decrease of the mRNA intensity logratio values (adenoma : normal tissue), respectively, from smaller to bigger adenomas *

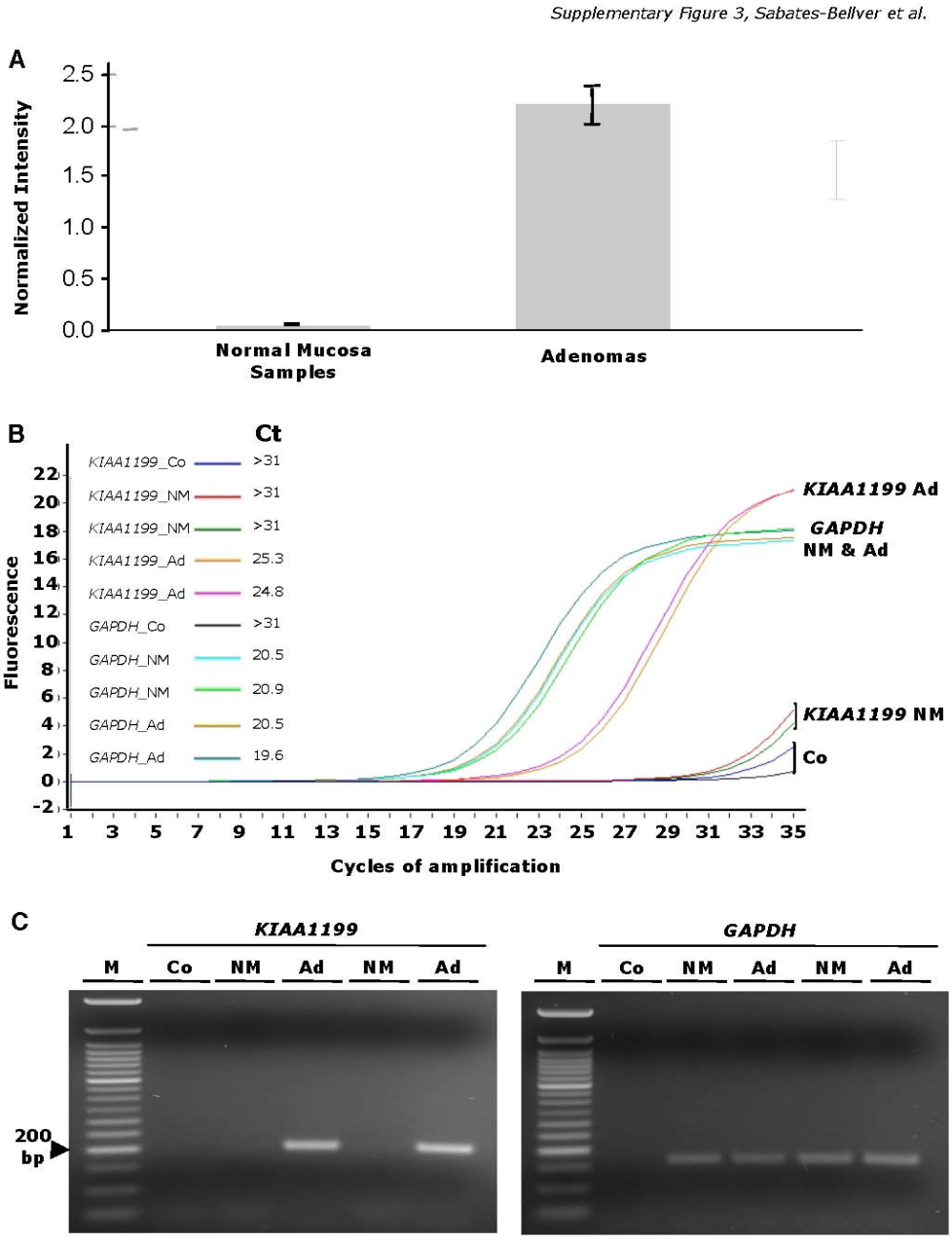
Positive correlation: Increasing fold changes with increasing adenoma size

Affymetrix ID	Gene Symbol	Gene Name
219189_at	**FBXL6**	F-box and leucine-rich repeat protein 6
235162_at	**MDM4**	Mdm4, transformed 3T3 cell double minute 4, p53 binding protein
32062_at	**LRRC14**	leucine rich repeat containing 14
200821_at	**LAMP2**	lysosomal-associated membrane protein 2
227080_at	**ZNF697**	zinc finger protein 697
206752_s_at	**DFFB**	DNA fragmentation factor, 40kDa, beta polypept. (caspase-activated DNase)
226334_s_at	**AHSA2**	AHA1, activator of heat shock 90kDa protein ATPase homolog 2
228150_at	**LZTR2**	leucine zipper transcription regulator 2
203140_at	**BCL6**	B-cell CLL/lymphoma 6 (zinc finger protein 51)
222557_at	**STMN3**	stathmin-like 3
226125_at	**SLC9A3**	solute carrier family 9 (sodium/hydrogen exchanger), member 3
201149_s_at	**TIMP3**	TIMP metallopeptidase inhibitor 3 (Sorsby fundus dystrophy)
1557580_at	**CABIN1**	Calcineurin binding protein 1
230097_at	**GART**	Phosphoribosylglycinamide formyltransferase
219359_at	**ATHL1**	ATH1, acid trehalase-like 1 (yeast)
1555259_at	**ZAK**	sterile alpha motif and leucine zipper containing kinase AZK
232201_at	**NKD2**	naked cuticle homolog 2 (Drosophila)
221499_s_at	**STX16**	syntaxin 16
224566_at	**TncRNA**	trophoblast-derived noncoding RNA
238449_at	**LOC440345**	hypothetical protein LOC440345
241815_at	**EST**	transcribed locus, hypothetical protein XP_498467 [Homo sapiens]
230494_at	**EST**	EST
229390_at	**RP1-93H18.5**	hypothetical protein LOC441168
232052_at	**LOC440944**	hypothetical gene supported by AK128398
1563088_a_at	**LOC284837**	hypothetical protein LOC284837
232052_at	**LOC440944**	Hypothetical gene supported by AK128398

Negative correlation: Decreasing fold changes with increasing adenoma size

216288_at	**CYSLTR1**	cysteinyl leukotriene receptor 1
37226_at	**BNIP1**	BCL2/adenovirus E1B 19kDa interacting protein 1
222377_at	**TBX10**	T-box 10
228962_at	**PDE4D**	phosphodiesterase 4D, cAMP-specific
229053_at	**SYT17**	synaptotagmin XVII
230360_at	**GLDN**	gliomedin
234709_at	**CAPN13**	calpain 13
239272_at	**MMP28**	matrix metallopeptidase 28
230360_at	**TMEM61**	transmembrane protein 61
223969_s_at	**RETNLB**	resistin like beta
212741_at	**MAOA**	monoamine oxidase A
240106_at	**GNPTAB**	N-acetylglucosamine-1-phosphate transferase, alpha and beta subunits
225822_at	**TMEM125**	transmembrane protein 125
213148_at	**LOC257407**	hypothetical protein LOC257407
229064_s_at	**EST**	EST

* These genes were selected from the list of top 500 probes with the highest loading scores along the CCA axis 1 (see Figure 2). Selection criteria were based on the statistical relation between expression change and adenoma size (i.e., how expression change related to adenoma size).



Supplementary Figure 3. ***KIAA1199*** **messenger RNA expression in colorectal adenomas.** (**A**) *KIAA1199* mRNA levels detected in samples of normal colorectal mucosa and adenomas with the Affymetrix U133 Plus 2.0 array based on normalized raw signal values (*y*-axis). Levels in adenomas were 54.8 times higher than those found in the normal mucosa. This difference was highly significant ($p < 0.0001$; bars: confidence intervals). (**B**) Real-time RT-PCR amplification of the coding region of *KIAA1199* and the control housekeeping gene *GAPDH* in two representative adenomas and matched samples of normal mucosa. (A total of 15 tissue pairs were examined in triplicate, including five collected after the present study was completed.) Ct (threshold cycle): number of cycles (*x*-axis) at the beginning of the log phase of amplification. After 35 cycles of PCR, *KIAA1199* was amplified in adenomas but not in the corresponding normal mucosa. At 40 cycles, a PCR product was also detected in many samples of normal mucosa (data not shown). (**C**) The *KIAA1199* amplification products were subjected to agarose-gel electrophoresis to verify the specificity of the amplification. M: 50-bp ladder marker; Co: negative control; NM: normal mucosa; Ad: adenoma.

Supplementary Table 4: Genes known to be targets of Wnt/β-catenin signaling #

Gene Symbol	Gene Name	Ref.
ABCB1 (MDR1)	ATP-binding cassette, sub-family B (MDR/TAP), member 1	*
APCDD1	adenomatosis polyposis coli down-regulated 1	1
ASCL2	achaete-scute complex-like 2 (Drosophila)	2
ATOH1 (Hath1)	Homo sapiens atonal homolog 1 (Drosophila) (ATOH1), mRNA.	*
AXIN2	axin 2 (conductin, axil)	*
BGLAP	bone gamma-carboxyglutamate (gla) protein (osteocalcin)	*
BIRC5	baculoviral IAP repeat-containing 5 (survivin)	*
BMP4	bone morphogenetic protein 4	*
BTRC	beta-transducin repeat containing	*
CCND1	cyclin D1 (PRAD1: parathyroid adenomatosis 1)	*
CD24	CD24 antigen (small cell lung carcinoma cluster 4 antigen)	3
CD44	CD44 antigen (homing function and Indian blood group system)	*
CDH1	cadherin 1, type 1, E-cadherin (epithelial)	*
CDX4	caudal type homeo box transcription factor 4	*
CLDN1	claudin 1	*
CSPG2	chondroitin sulfate proteoglycan 2 (versican)	*
CST1	cystatin SN	4
CTGF	connective tissue growth factor	*
CYR61	cysteine-rich, angiogenic inducer, 61	*
DKK1	dickkopf homolog 1 (Xenopus laevis)	*
DLL1	delta-like 1 (Drosophila)	*
DRCTNNB1A	down-regulated by Ctnnb1, a	5
EDA	ectodysplasin A	*
EDAR	ectodysplasin 1, anhidrotic receptor	*
EDN1	endothelin 1	*
EDN3	endothelin 3	4
EFNB1	ephrin-B1	*
EGFR	epidermal growth factor receptor	*
EMP1	epithelial membrane protein 1	*
EN2	engrailed homolog 2	*
ENPP2	ectonucleotide pyrophosphatase/phosphodiesterase 2 (autotaxin)	*
EPHB2	EPH receptor B2	*
EPHB3	EPH receptor B3	*
ETV4	ets variant gene 4 (E1A enhancer binding protein, E1AF)	6
FGF18	fibroblast growth factor 18	*
FGF20	fibroblast growth factor 20	*
FGF4	fibroblast growth factor 4	*
FGF9	fibroblast growth factor 9 (glia-activating factor)	*
FN1	fibronectin 1	*
FOSL1 (fra-1)	FOS-like antigen 1	*
FOXN1	forkhead box N1	*
FST	follistatin	*
FZD7	frizzled homolog 7 (Drosophila)	*
GAS	gastrin	*
GCG	glucagon	*
GJA1	gap junction protein, alpha 1, 43kDa (connexin 43)	*
GJB6	gap junction protein, beta 6 (connexin 30)	*
HIG2	hypoxia-inducible protein 2	7

Supplementary Table 4, page 2

HSPG2	heparan sulfate proteoglycan 2 (perlecan)	*
ID2	inhibitor of DNA binding 2, dominant negative helix-loop-helix protein	*
IGF1	insulin-like growth factor 1 (somatomedin C)	*
IGF2	insulin-like growth factor 2 (somatomedin A)	*
IGFBP2	insulin-like growth factor binding protein 2, 36kDa	8
IL6	interleukin 6 (interferon, beta 2)	*
IL8	interleukin 8	*
IRX3	iroquois homeobox protein 3	*
ISLR	immunoglobulin superfamily containing leucine-rich repeat	*
JUN	v-jun sarcoma virus 17 oncogene homolog (avian)	*
KLF5 (BTEB2)	Kruppel-like factor 5 (intestinal)	*
KRTHB6	keratin, hair, basic, 6 (monilethrix)	*
L1CAM	L1 cell adhesion molecule	*
LEF1	lymphoid enhancer-binding factor 1	*
LGR5	leucine-rich repeat-containing G protein-coupled receptor 5	9
LRP5	low density lipoprotein receptor-related protein 5	*
LRP6	low density lipoprotein receptor-related protein 6	*
MET	met proto-oncogene (hepatocyte growth factor receptor)	*
MITF	microphthalmia-associated transcription factor	*
MMP26	matrix metalloproteinase 26	*
MMP3	matrix metalloproteinase 3 (stromelysin 1, progelatinase)	*
MMP7	matrix metalloproteinase 7 (matrilysin, uterine)	*
MSLN	mesothelin	*
MSX1	msh homeo box homolog 1 (Drosophila)	10
MSX2	msh homeo box homolog 2 (Drosophila)	10
MYC	v-myc myelocytomatosis viral oncogene homolog (avian)	*
MYCBP	c-myc binding protein	*
NEUROG1	neurogenin 1	*
NKD1	naked cuticle homolog 1 (Drosophila)	*
NKD2	naked cuticle homolog 2 (Drosophila)	*
NLK	nemo like kinase	*
NRCAM	neuronal cell adhesion molecule	*
OVOL1 (movo1)	ovo-like 1(Drosophila)	*
PITX2	paired-like homeodomain transcription factor 2	*
PLAUR (uPar)	plasminogen activator, urokinase receptor	*
POSTN	periostin, osteoblast specific factor	*
PPARD (PPARd)	peroxisome proliferative activated receptor, delta	*
PTGS2 (COX-2)	prostaglandin-endoperoxide synthase 2	*
PTTG1	pituitary tumor-transforming 1	*
RARG	retinoic acid receptor, gamma	*
RET	ret proto-oncogene	*
RHOU (Wrch-1)	ras homolog gene family, member U	*
RUNX2	runt-related transcription factor 2	*
SFRP2	secreted frizzled-related protein 2	*
SIX3	sine oculis homeobox homolog 3 (Drosophila)	*
SOX2	SRY (sex determining region Y)-box 2	*
SOX4	SRY (sex determining region Y)-box 4	11
SOX9	SRY (sex determining region Y)-box 9	*
SP5	Sp5 transcription factor	12

Supplementary Table 4, page 3

STRA6	stimulated by retinoic acid gene 6	*
T (Brachyury)	T, brachyury homolog (mouse)	*
TCF1	transcription factor 1, hepatic; LF-B1, hepatic nuclear factor (HNF1)	*
TCF4 (ITF2)	transcription factor 4	*
TIAM1	T-cell lymphoma invasion and metastasis 1	*
TNFSF11 (RANKL)	tumor necrosis factor (ligand) superfamily, member 11	*
TNFSF9 (4-1BBL)	tumor necrosis factor (ligand) superfamily, member 9	*
TWIST1	twist homolog 1 (acrocephalosyndactyly 3; Saethre-Chotzen syndrome)	*
UBXD1	UBX domain containing 1	*
VEGF	vascular endothelial growth factor	*
VEGFC	vascular endothelial growth factor C	*
WISP1	WNT1 inducible signaling pathway protein 1	*
WISP2	WNT1 inducible signaling pathway protein 2	*

This gene list is based on the data of R. Nusse (http://www.stanford.edu/~rnusse/pathways/targets.html and http://www.stanford.edu/~rnusse/pathways/targetcomp.html (studies referenced with *) and those reported in the references listed below. The is not intended to be comprehensive and does not include genes identified as putative Wnt targets in gene expression studies.

1. Cancer Res. 2002;**62**:5651-5656.
2. Oncogene. 2006;**25**:3445-3457.
3. Oncogene. 2006;**25**:4361-4369.
4. Cancer Res. 2003;**63**:2913-2922.
5. Cancer Res. 2000;**60**:3354-3358
6. Am J Pathol. 2002;**160**:2181-2190.
7. BMC Cancer. 2005;**5**:3.
8. Oncogene. 2005;**24**:3141-3153.
9. Hepatology. 2003;**37**:528-533.
10. BMC Dev Biol. 2002;**2**:8.
11. Cancer Res. 2005;**65**:166-176.
12. Int J Oncol. 2005;**27**:1483-1487.

Supplementary Figure 4, Sabates-Bellver et al.

A

		Known Wnt target genes		
		yes	no	Total
Genes correlated	yes	**19**	338	357
with *KIAA1199*	no	20	3930	3950
	Total	39	4268	4307

Fisher's exact test (p-value= 2.6e-11)

B

The 19 known Wnt target genes whose expression was correlated with that of *KIAA1199*

Affymetrix ID	Gene Symbol	PC* value
222696_at	***AXIN2***	0.95
209835_x_at	***CD44***	0.92
203510_at	***MET***	0.91
202431_s_at	***MYC***	0.90
202936_s_at	***SOX9***	0.88
229215_at	***ASCL2***	0.88
204259_at	***MMP7***	0.87
222549_at	***CLDN1***	0.86
218507_at	***HIG2***	0.84
209589_s_at	***EPHB2***	0.81
205828_at	***MMP3***	0.81
211506_s_at	***IL8***	0.81
1438_at	***EPHB3***	0.81
202095_s_at	***BIRC5***	0.80
210319_x_at	***MSX2***	0.80
213880_at	***LGR5***	0.80
201417_at	***SOX4***	0.80
202718_at	***IGFBP2***	0.80
205932_s_at	***MSX1***	0.80

* PC, Pearson Correlation

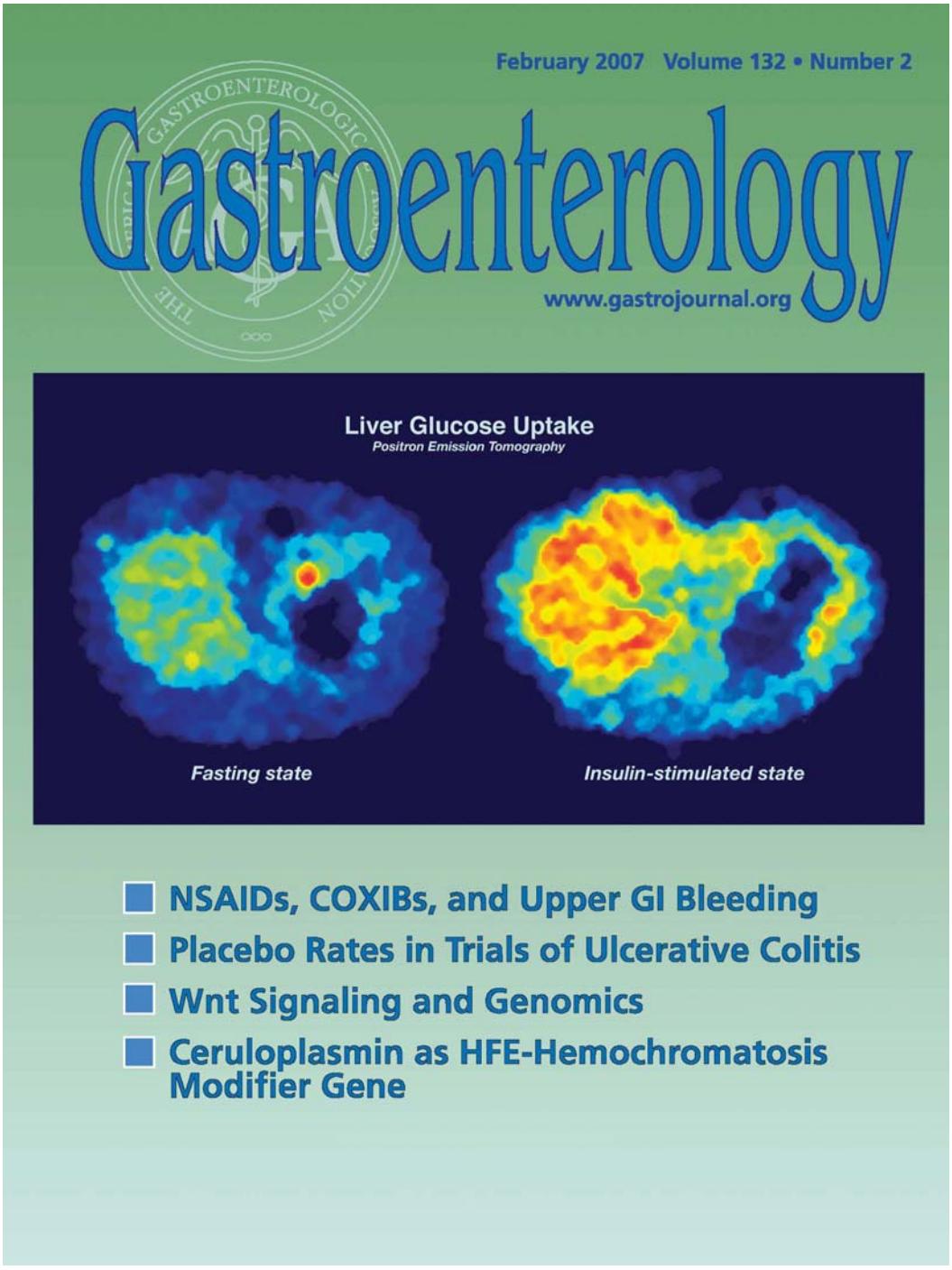
Supplementary Figure 4. **Degree of co-expression between *KIAA1199* and known Wnt target genes.** (**A**) 4307 genes (6379 probes) were significantly up-regulated in adenomas, compared with normal mucosa (no fold-change restriction). This set included 39 of the 110 known Wnt target genes listed in Supplementary Table 4 and 357 genes whose expression appeared to be co-regulated with that of *KIAA1199* (Pearson correlation value of ≥0.8). Comparison of these two subsets revealed that 49% (19/39) of the known Wnt target genes that were significantly over-expressed in our adenomas also displayed positive correlation with *KIAA1199* expression (panel **B** and **Figure 3A and 3B**). In contrast, co-regulation with *KIAA1199* was much less common among the upregulated genes that are not currently considered to be Wnt targets (338/4268, 7.9%) (Fisher exact test; $p<0.0001$).

Supplementary Table 5, page 2

204470_at	6.6	CXCL1	NM_001511	chemokine (C-X-C motif) ligand 1 (melanoma growth stimulating activity, alpha)	chemotaxis / inflammatory response / G-protein coupled receptor protein signaling pathway / intracellular signaling cascade / nervous system development / negative regulation of cell proliferation / actin cytoskeleton organization and biogenesis / immune response / signal transduction
229613_at	6.5		BF724178	Weakly similar to naked cuticle homolog 1; naked cuticle-1; Dvl-binding protein [Pan troglodytes]	unknown
1568603_at	6.4	CADPS	AI912173	Ca2+-dependent secretion activator	exocytosis / protein transport
202376_at	6.4	SERPINA3	NM_001085	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 3	acute-phase response / inflammatory response / regulation of lipid metabolism
226034_at	6.2		BE222344	Homo sapiens, clone IMAGE:3881549, mRNA	unknown
219000_s_at	6.0	PHLDA1	NM_007350	pleckstrin homology-like domain, family A, member 1	apoptosis
1561330_at	5.9	DSG4	BC039098	desmoglein 4	cell adhesion
207033_at	5.9	GIF	NM_005142	gastric intrinsic factor (vitamin B synthesis)	cobalt ion transport
227051_at	5.8		AU157716	Homo sapiens, clone IMAGE:4393354, mRNA	unknown
213880_at	5.8	LGR5	AL524520	leucine-rich repeat-containing G protein-coupled receptor 5	signal transduction / G-protein coupled receptor protein signaling pathway
226754_at	5.8	SLC6A6	BG160485	solute carrier family 6 (neurotransmitter transporter, taurine), member 6	amino acid metabolism / neurotransmitter transport
232203_at	5.7		AA654714	cDNA FLJ13722 fis, clone PLACE2000455	unknown
212531_at	5.7	LCN2	NM_005564	lipocalin 2 (oncogene 24p3)	transport
228303_at	5.6		AW014155	Similar to UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 6 (GalNAc-T6)	unknown
219956_at	5.6	GALNT6	NM_007210	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase	protein amino acid O-linked glycosylation
211207_s_at	5.6	ACSL6	AF129168	acyl-CoA synthetase long-chain family member 6	lipid metabolism / acyl-CoA metabolism / fatty acid metabolism
216325_x_at	5.6	RTEL1	AF217796	regulator of telomere elongation helicase 1	nucleobase, nucleoside, nucleotide and nucleic acid metabolism
228955_at	5.6		AU041761	Weakly similar to NP_990560.1 very low density lipoprotein (VLDL)/vitellogenin receptor [Gallus gallus]	unknown
206925_at	5.5	PCSK1	NM_000439	proprotein convertase subtilisin/kexin type 1	proteolysis / cell-cell signaling / metabolism
225541_at	5.5	RPL22L1	BE274422	ribosomal protein L22-like 1	protein biosynthesis / electron transport
215185_at	5.5		AK024177	Homo sapiens, clone IMAGE:3851018, mRNA	unknown
1554676_a_at	5.4	ETV4	BC007242	ets variant gene 4 (E1A enhancer binding protein, E1AF)	regulation of transcription, DNA-dependent
236910_s_at	5.4	LRRC34	AW975772	leucine rich repeat containing 34	unknown
206172_at	5.4	IL13RA2	NM_000640	interleukin 13 receptor, alpha 2	unknown
220003_at	5.3	LRRC36	NM_018096	leucine rich repeat containing 36	unknown
220668_s_at	5.3	ARNTL2	NM_020183	aryl hydrocarbon receptor nuclear translocator-like 2	regulation of transcription, DNA-dependent / signal transduction // entrainment of circadian clock / regulation of transcription from RNA polymerase II promoter / protein modification / proteolysis
205174_s_at	5.3	QPCT	NM_012413	glutaminyl-peptide cyclotransferase (glutaminyl cyclase)	unknown
223694_at	5.3	TRIM7	AF220032	tripartite motif-containing 7	skeletal development / regulation of transcription, DNA-dependent / regulation of transcription from RNA polymerase II promoter / muscle development
204201_at	5.2	TEAD4	NM_003213	TEA domain family member 4	organic cation transport
205421_at	5.2	SLC22A3	NM_021977	solute carrier family 22 (extraneuronal monoamine transporter), member 3	xenobiotic metabolism
232654_s_at	5.2	UGT1A6	BE873420	UDP glucuronosyltransferase 1 family, polypeptide A6	cell adhesion / visual perception / cell proliferation / response to stimulus
201506_at	5.2	TGFBI	NM_000358	transforming growth factor, beta-induced, 68kDa	regulation of transcription, DNA-dependent
236945_at	5.2	SP5	AI360207	Sp5 transcription factor	unknown
244612_at	5.2		AW117101	EST	unknown
1568639_a_at	5.1		BC033867	Homo sapiens, clone IMAGE:5110701, mRNA	unknown
228849_at	5.1	NOTUM	AL574571	notum pectinacetylesterase homolog (Drosophila)	monocarboxylic acid transport
205234_at	5.0	SLC16A4	NM_004696	solute carrier family 16 (monocarboxylic acid transporters), member 4	chemotaxis / inflammatory response / G-protein coupled receptor protein signaling pathway / response to stimulus
209774_x_at	5.0	CXCL2	M57731	chemokine (C-X-C motif) ligand 2	complement activation, classical pathway / blood coagulation / innate immune response
208209_s_at	4.8	C4BPB	NM_000716	complement component 4 binding protein, beta	regulation of cell growth
202718_at	4.8	IGFBP2	NM_000597	insulin-like growth factor binding protein 2, 36kDa	intracellular signaling cascade
230398_at	4.8	TNS4	AA158731	tensin 4	calcium ion homeostasis / cell surface receptor linked signal transduction / cell-cell signaling / response to nutrient
230746_s_at	4.7	STC1	AW003173	Stanniocalcin 1	signal transduction / G-protein coupled receptor protein signaling pathway / negative regulation of adenylate cyclase activity / synaptic transmission / visual perception / sensory perception of smell / response to stimulus
216255_s_at	4.7	GRM8	AC000099	glutamate receptor, metabotropic 8	activation of MAPK activity / morphogenesis of a branching structure / epidermal growth factor receptor signaling pathway / heart development / positive regulation of cell proliferation / determination of anterior/posterior axis, embryo / regulation of signal transduction / peptidyl-serine phosphorylation / cell differentiation / positive regulation of cell migration / mammary gland development / negative regulation of apoptosis / positive regulation of peptidyl-tyrosine phosphorylation / mesodermal cell fate determination
206206_s_at	4.7	TDGF1	NM_003212	teratocarcinoma-derived growth factor 1	cell adhesion
204905_s_at	4.7	MSLN	NM_005823	mesothelin	regulation of smooth muscle contraction / signal transduction / neuropeptide signaling pathway / digestion / G-protein coupled receptor protein signaling pathway
206023_at	4.7	NMU	NM_006681	neuromedin U	regulation of transcription, DNA-dependent / cell differentiation / development
226445_at	4.6	HES6	AW249678	hairy and enhancer of split 6 (Drosophila)	cell adhesion / negative regulation of cell proliferation / lipid catabolism / antigen presentation / cell communication
209309_at	4.6	AZGP1	D90427	alpha-2-glycoprotein 1, zinc	skeletal development / regulation of transcription from RNA polymerase II promoter / morphogenesis / regulation of transcription, DNA-dependent
229576_s_at	4.6	TBX3	N29712	T-box 3 (ulnar mammary syndrome)	unknown
219270_at	4.5	CHAC1	NM_024111	ChaC, cation transport regulator-like 1 (E. coli)	nucleobase, nucleoside, nucleotide and nucleic acid metabolism / apoptosis
206467_x_at	4.5	TNFRSF6B	NM_003823	tumor necrosis factor receptor superfamily, member 6b, decoy	endothelial cell migration
204351_at	4.5	S100P	NM_005980	S100 calcium binding protein P	immune response / negative regulation of angiogenesis / cytokine and chemokine mediated signaling pathway / platelet activation / immune cell chemotaxis / negative regulation of megakaryocyte differentiation / response to stimulus / chemotaxis
206390_x_at	4.4	PF4	NM_002619	platelet factor 4 (chemokine (C-X-C motif) ligand 4)	unknown
225905_at	4.4	JUB	AI289311	jub, ajuba homolog (Xenopus laevis)	DNA replication / deoxyribonucleoside diphosphate metabolism
209773_s_at	4.4	RRM2	BC001886	ribonucleotide reductase M2 polypeptide	unknown
235153_at	4.4	RNF183	BE796148	ring finger protein 183	cell adhesion
212014_x_at	4.4	CD44	AI493245	CD44 molecule (Indian blood group)	intracellular protein transport / endocytosis / vesicle-mediated transport
1555731_a_at	4.4	AP1S3	AF393369	adaptor-related protein complex 1, sigma 3 subunit	amino acid metabolism / amino acid transport
201195_s_at	4.3	SLC7A5	AB018009	solute carrier family 7 (cationic amino acid transporter, y+ system), member 5	regulation of transcription, DNA-dependent / transcription from RNA polymerase II promoter
236471_at	4.3	NFE2L3	AI949827	nuclear factor (erythroid-derived 2)-like 3	unknown
223974_at	4.3	MGC11082	BC005130	hypothetical protein MGC11082	unknown
219710_at	4.3	SH3TC2	NM_024577	SH3 domain and tetratricopeptide repeats 2	electron transport / xenobiotic metabolism / nitric oxide biosynthesis / synaptic transmission, cholinergic / response to toxin
201467_s_at	4.2	NQO1	AI039874	NAD(P)H dehydrogenase, quinone 1	chemotaxis / inflammatory response / signal transduction / cell-cell signaling / antimicrobial humoral response / defense response to bacteria / response to stimulus / chemotaxis / immune response
205476_at	4.2	CCL20	NM_004591	chemokine (C-C motif) ligand 20	unknown
227919_at	4.2	UCA1	AA702248	UCA1 protein	inflammatory response
202917_s_at	4.2	S100A8	NM_002964	S100 calcium binding protein A8 (calgranulin A)	unknown
222830_at	4.2	GRHL1	BE566136	grainyhead-like 1 (Drosophila)	cell cycle / epithelial cell differentiation / induction of bacterial agglutination / innate immune response / negative regulation of progression through cell cycle
208250_s_at	4.2	DMBT1	NM_004406	deleted in malignant brain tumors 1	cell communication / cell adhesion / development
230964_at	4.1	FREM2	N66307	FRAS1 related extracellular matrix protein 2	regulation of progression through cell cycle / carbohydrate metabolism / base-excision repair / DNA repair / response to DNA damage stimulus
210021_s_at	4.1	UNG2	BC004077	uracil-DNA glycosylase 2	signal transduction / G-protein coupled receptor protein signaling pathway
205240_at	4.1	GPSM2	NM_013296	G-protein signalling modulator 2 (AGS3-like, C. elegans)	unknown
225767_at	4.1		AI829933	EST, weakly similar to PIHUD6 salivary proline-rich protein precursor PRB1	apoptosis / signal transduction
211163_s_at	4.1	TNFRSF10C	AF012536	tumor necrosis factor receptor superfamily, member 10c, decoy without an intracellular domain	

6.2. Appendix II

"The intestinal Wnt/TCF signature" Laurens G. Van der Flier*, Jacob Sabates-Bellver*, Irma Oving, Andrea Haegebarth, Mariagrazia de Palo, Marcello Anti, Mareielle E. van Gijn, Saskia Suijkerbuijk, Marc Van de Wetering, Giancarlo Marra and Hans Clevers. Gastroenterology 2007;132:628-32. *These authors contribute equally to this work



GASTROENTEROLOGY 2007;132:628–632

The Intestinal Wnt/TCF Signature

LAURENS G. VAN DER FLIER,* JACOB SABATES-BELLVER,‡ IRMA OVING,* ANDREA HAEGEBARTH,* MARIAGRAZIA DE PALO,§ MARCELLO ANTI,§ MARIELLE E. VAN GIJN,* SASKIA SUIJKERBUIJK,* MARC VAN DE WETERING,* GIANCARLO MARRA,‡ and HANS CLEVERS*

*Hubrecht Institute, Netherlands Institute for Developmental Biology and Centre for Biomedical Genetics, Utrecht, The Netherlands; ‡Institute of Molecular Cancer Research, University of Zurich, Zurich, Switzerland; and §the Gastroenterology Unit, Belcolle Hospital, Strada Sammartinese, Viterbo, Italy

Background & Aims: **In colorectal cancer, activating mutations in the Wnt pathway transform epithelial cells through the inappropriate expression of a TCF4 target gene program, which is physiologically expressed in intestinal crypts.** ***Methods:*** **We have now performed an exhaustive array-based analysis of this target gene program in colorectal cancer cell lines carrying an inducible block of the Wnt cascade. Independently, differential gene-expression profiles of human adenomas and adenocarcinomas vs normal colonic epithelium were obtained.** ***Results:*** **Expression analyses of approximately 80 genes common between these data sets were performed in a murine adenoma model. The combined data sets describe a core target gene program, the intestinal Wnt/TCF signature gene set, which is responsible for the transformation of human intestinal epithelial cells.** ***Conclusions:*** **The genes were invariably expressed in adenomas, yet could be subdivided into 3 modules, based on expression in distinct crypt compartments. A module of 17 genes was specifically expressed at the position of the crypt stem cell.**

The overwhelming majority of colorectal cancers (CRCs) is initiated by activating mutations in the Wnt pathway.[1,2] These mutations either remove the tumor suppressors adenomatous polyposis coli (APC), or axin or activate the proto-oncogene β-catenin. As a common result, β-catenin accumulates in the nucleus, and constitutively binds to the T-cell factor (TCF)4 transcription factor, resulting in transcriptional activation of Wnt/TCF4 target genes, initiating transformation of intestinal epithelial cells.[3,4] Physiologically, the Wnt pathway is essential for the maintenance of crypt progenitor compartments, as evidenced by mice lacking the TCF4 transcription factor,[5] or transgenically expressing the secreted Dickkopf-1 Wnt inhibitor.[6,7]

A multitude of reports have appeared in which single-candidate TCF4 target genes have been described (eg, references listed in Table 1 and available at http://www.stanford.edu/~rnusse/pathways/targets.html). Typically, these studies describe differential expression between cells with and without an activated Wnt pathway, followed by transient promoter assays. DNA array technology allows the assessment of differential messenger RNA (mRNA) expression on a genome-wide scale.

Materials and Methods

Cell Culture

CRC cell lines LS174T and DLD1, stably expressing inducible dominant-negative (dn)TCF1 or dnTCF4, were generated as previously described.[8] The Wnt pathway activity in the CRC cells was determined as described previously[3] using the optimized TCF reporter pTopGlow and its negative control pFopGlow, constructed in our laboratory.[9]

Oligonucleotide Microarray Analysis of CRC Cell Lines

RNA was isolated after 10 and 20 hours induction of the dnTCFs. RNA quality was assessed using capillary gel electrophoresis (BioAnalyzer; Agilent Technologies). Complementary DNA (cDNA) synthesis and labeling was performed according to Affymetrix (Santa Clara, CA) guidelines. cRNA was synthesized and labeled with Affymetrix One-cycle Target Labeling kit, and hybridized on Affymetrix GeneChip HG-U133 plus 2.0 microarrays. The overall fluorescence for each GeneChip was scaled to a target intensity of 200. The expression profiles at 10 and 20 hours after induction were compared with those of noninduced controls by pair-wise comparisons performed with GeneChip Operating Software (Affymetrix). Only probes with a significantly decreased call for both time points were included.

Oligonucleotide Microarray Analysis of Human Tissues

Colorectal adenomas were collected during endoscopy at a single gastroenterology unit (Belcolle Hospital, Viterbo, Italy) with full institutional review board approval. In each patient, normal mucosa also was collected at about a 2–4-cm distance from the adenoma. Biopsy specimens

Abbreviations used in this paper: APC, adenomatous polyposis coli; CRC, colorectal cancer; dn, dominant-negative; TCF, T-cell factor; wt, wild type.

0016-5085/07/$32.00
doi:10.1053/j.gastro.2006.08.039

Table 1. Target Genes Down-Regulated in All 4 CRC Cell Lines on Over Expression of dnTCFs

			LS174T dnTCF1		LS174T dnTCF4		DLD1 dnTCF1		DLD1	
Gene symbol	References	Affymetrix ID	10 h	20 h	10 h	20 h	10 h	20 h	10 h	20 h
ASCL2	18	229215_at	−4.3	−24	−2.3	−15	−2.8	−7.5	−3.0	−4.0
AXIN2	14–16	222696_at	−2.5	−4.9	−2.3	−2.6	−2.8	−3.5	−2.8	−3.7
BMP4[a]	21	211518_s_at	−2.0	−2.5	−2.3	−4.9	−3.0	−3.2	−1.6	−2.3
C1orf33[a]		220688_s_at	−1.7	−3.2	−2.3	−2.8	−1.7	−2.0	−1.6	−2.0
HIG2	20	1554452_a_at	−1.7	−4.0	−2.1	−3.2	−1.4	−2.6	−1.9	−2.0
HSPC111		203023_at	−1.5	−3.0	−2.3	−2.5	−1.4	−1.6	−1.9	−1.7
HSPC111		214011_s_at	−1.6	−3.0	−2.3	−2.3	−1.5	−1.9	−1.6	−1.6
KITLG[a]		226534_at	−2.6	−3.0	−2.3	−2.3	−4.3	−2.6	−1.6	−2.8
LGR5[a]	19	213880_at	−4.3	−9.8	−2.3	−3.5	−7.5	−7.5	−2.1	−3.2
MYC[a]	17	202431_s_at	−2.1	−2.3	−2.3	−1.6	−2.8	−3.0	−2.1	−2.3
NOL1		214427_at	−1.7	−2.8	−2.3	−2.1	−1.3	−1.9	−1.6	−1.6
PPIF		201490_s_at	−1.4	−1.6	−2.3	−2.0	−1.5	−1.9	−1.5	−1.5
SOX4	22	201416_at	−1.3	−2.1	−2.3	−2.1	−3.0	−3.2	−2.1	−2.0
WDR71		218957_s_at	−2.8	−17	−2.3	−2.8	−1.9	−3.7	−2.0	−3.7
ZIC2		223642_at	−1.7	−2.3	−2.3	−2.3	−1.4	−2.1	−1.7	−1.9
ZNRF3[a]		226360_at	−4.6	−3.2	−2.3	−2.5	−3.5	−3.2	−2.1	−2.8

NOTE. Fold changes of genes down-regulated in all 4 CRC cell lines on over expression of dnTCFs are shown.
[a]Genes also have been identified in the Stanford array experiment.[8]

were immersed immediately in RNAlater (Ambion, Huntingdon, UK), homogenized, and RNA was extracted with RNeasy kit (QIAgen, Basel, Switzerland). RNA quality was verified by capillary gel electrophoresis and a total of 32 pairs of normal mucosa and adenomas was analyzed. cRNA was synthesized, labeled, and hybridized as described earlier. GeneSpring software (Silicon Genetics, Redwood City, CA) was used for gene expression data and statistical analyses. The nonparametric Mann–Whitney test was used with a false discovery rate of .05 and Bonferroni correction for the group comparison analysis (normal mucosa vs adenomas or cancer). Transcriptome data from a series of 25 previously collected colon adenocarcinomas and 10 samples of normal mucosa from some of these patients were obtained with the same procedure[10] and used in this study.

In Situ Hybridizations

Mouse orthologs of selected transcripts were obtained as expressed sequence tags IMAGE consortium (MRC geneservise, Badrahman, UK) or RZPD (German Resource Center for Genome Research, Berlin, Germany). These clones were used for in vitro transcription reactions to generate probes for in situ hybridizations. Protocols for in vitro transcription and in situ hybridizations are described elsewhere.[11]

Results

Inhibition of the Constitutively Active Wnt Pathway in CRC Cells

N-terminally truncated TCFs do not bind β-catenin and act as potent inhibitors of endogenous β-catenin/TCF complexes.[12] Tcf4 is expressed physiologically in the intestine.[13] In our hands, the DNA-binding characteristics of TCF4 and TCF1 are essentially identical. For TCF target gene identification, we generated a panel of cell clones from the CRC cell lines LS174T (mutationally activated allele of the *CTNNB1* gene encoding β-catenin) and in DLD1 (mutant APC), in which the Wnt cascade could be inhibited by inducible expression of either dnTCF1 or dnTCF4. Individ-

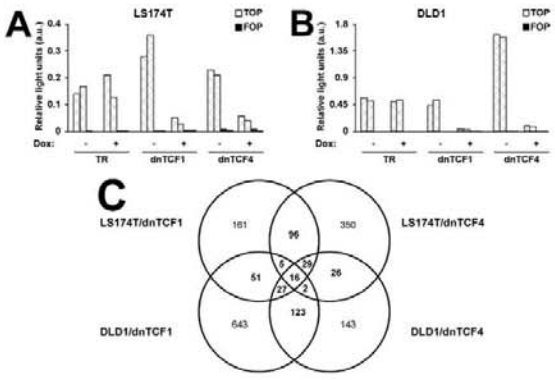


Figure 1. Over expression of dnTCFs inhibits the Wnt pathway in CRC cells. TCF/β-catenin driven transcription is abrogated by over expression of dnTCF1 or dnTCF4 in (*A*) LS174T cells and (*B*) DLD1 cells. The activity of the TCF reporter, pTopGlow (□, TOP), and the control, pFopGlow (■, FOP), after 20 hours with or without doxycycline treatment is shown. Parental cells that expressed the tetracycline repressor (TR) were used as controls. Renilla luciferase levels were used as transfection controls. (*C*) Venn diagram showing the (overlapping) down-regulated probes on dnTCF1 or dnTCF4 induction. Selection is based on down-regulation in 2 or more of the 4 cell-line transfectants used in this study. Not shown in the figure is the overlap in 6 unique probes between LS174T/dnTCF1 and DLD1/dnTCF4 and 6 probes between LS174T/dnTCF4 and DLD1/dnTCF1.

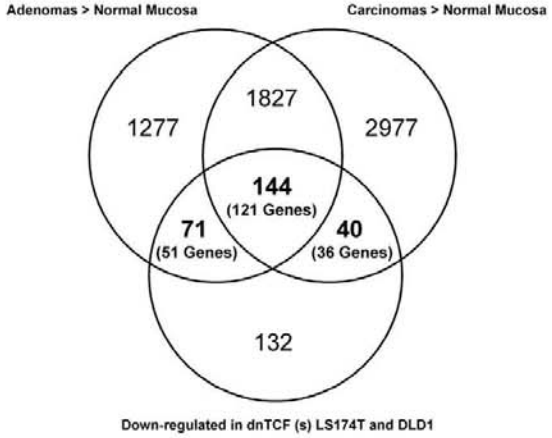


Figure 2. The intestinal Wnt/TCF signature gene set. TCF/β-catenin target genes were selected by comparing the 387 probes (lower circle) obtained with the procedure described in Figure 1*C* with the 8307 probes whose expression levels in human tumors were increased significantly, relative to those in normal mucosa. The left upper circle represents 3319 probes increased in adenomas, whereas the right upper circle represents 4988 probes increased in carcinomas (1971 were increased in both tumor types). With this approach, we classified 51 genes (71 probes) up-regulated in adenomas, 36 genes (40 probes) up-regulated in carcinomas, and 121 genes (144 probes) up-regulated in both tumor types as β-catenin/TCF target genes. These gene lists are reported in Supplementary Table 1 (supplementary material online at www.gastrojournal.org).

ual transfectants were selected based on the induced inhibition of the constitutive Wnt activity. For the 4 selected clones, the constitutively active Wnt pathway could be inhibited close to background levels on doxcycycline-mediated induction of dnTCF1 or dnTCF4 (Figure 1*A* and *B*). Moreover, all clones underwent a rapid, robust G1 arrest within 24 hours (not shown).

The Genetic Program Driven by β-Catenin/TCF in CRC Cell Lines

We originally subjected the LS174T/dnTCF4 clone to a microarray experiment on the Stanford spotted-cDNA array platform containing 24,000 probes.[8] To obtain a more comprehensive list of TCF target genes in intestinal cancer cells, we performed expression profiling using the Affymetrix GeneChip HG-U133 plus 2.0, which contains 54,675 probes. We initiated analyses with the LS174T/dnTCF4 cells. mRNA was isolated at 10 and 20 hours after induction and from noninduced control cells. The 115 probes reported on the Stanford platform[8] represented 101 significantly down-regulated genes, all of which also were present on the Affymetrix arrays. More than 60% of these 101 genes turned out to be down-regulated significantly at either the 10- or 20-hour time points in the Affymetrix measurements, whereas 26% were significantly down-regulated at both time points. This showed good reproducibility and robustness of the assay, given the extensive differences in technology between the 2 platforms.

Similar array experiments then were performed for the other 3 transfectants (LS174T/dnTCF1, DLD1/dnTCF4, and DLD1/dnTCF1). For each cell line, we compiled lists of probe features significantly down-regulated at both time points. Figure 1*C* gives a Venn diagram summarizing the down-regulated features shared between the 4 different cell

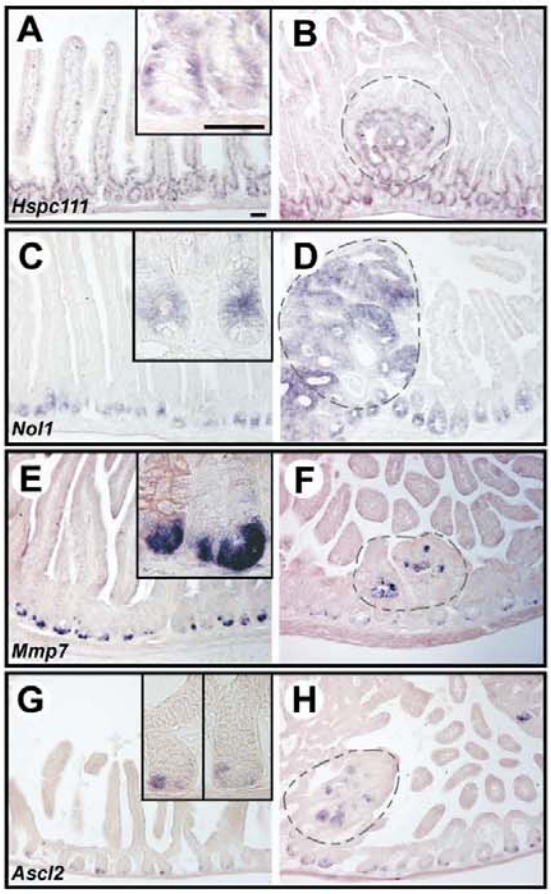


Figure 3. Confirmation and classification of TCF target genes by in situ hybridization on intestines of adult *APC*^{min} mice. All Wnt signature genes tested were expressed in (*A*, *C*, *E*, and *G*) wild type crypts and in (*B*, *D*, *F*, and *H*) adenomas of *Apc*^{min} mice. The staining patterns in wt crypts could be divided into 3 categories. (*A* and *B*) An example of the expression of a gene within the first category, the mouse ortholog of *HSPC111*, is shown. In addition to expression within the proliferative compartment of crypts and adenomas, a decreasing gradient of signal is observed along the base of the villus. (*C* and *D*) Another gene in this category, the mouse ortholog for *NOL1*, showed expression restricted to the proliferative compartment of crypts and adenomas, but lacked significant expression along the villus. (*E* and *F*) The second staining pattern involves the Paneth cells at the bottom of the crypts, as represented by the mouse ortholog for *MMP7*. (*G* and *H*) In situ hybridizations for the mouse ortholog of *ASCL2* defined a third category that revealed expression in adenomas and in a few cells near the crypt bottom. This position coincides with the location of the elusive crypt stem cells. Size bars represent 50 μm, and adenomas are encircled by dotted lines.

lines. The 15 genes down-regulated in all 4 cell lines are reported in Table 1. Table 1 includes some well-known Wnt targets *AXIN2*,[14–16] *c-Myc*,[17] *ASCL2*,[18] *LGR5*,[19] *HIG2*,[20] *BMP4*,[21] and *SOX4*,[22] as well as 8 novel TCF target genes.

To obtain an information-rich, yet robust data set, we arbitrarily selected the probes that were down-regulated at both time points in at least 2 different cell lines. This resulted in a list that contained 387 probe features. This list was used for comparison with the gene-expression profiles of adenomas and adenocarcinomas described later.

Gene Expression Profiles of Human Adenomas and Adenocarcinomas

The transcriptome of 32 adenomas and 25 carcinomas was analyzed and compared with that of normal colonic mucosa. The genes up-regulated in adenomas and/or carcinomas were compared with those of the 387 β-catenin/TCF responsive genes identified in the dnTCF CRC cell lines. Figure 2 shows the Venn diagram procedure we used to select the Wnt/TCF signature gene set. A total of 255 common probes corresponding to 208 different genes are reported in Supplemental Table 1 (supplementary material online at www.gastrojournal.org). A total of 121 of the genes were up-regulated in both tumor types, whereas 51 and 36 were up-regulated only in the adenomas or in carcinomas, respectively.

Confirmation of Target Genes by In Situ Hybridizations

In our initial study,[8] we observed that, as a rule, the identified TCF target genes were expressed physiologically in rapidly dividing crypt cells. It since has been observed that a subset of TCF target genes is expressed in the postmitotic Paneth cells, which are located at the crypt bottom.[23,24] We extended these observations for the intestinal Wnt/TCF signature gene set by in situ hybridizations on intestinal tissue derived from adult mice carrying the Apc^{min} allele.[25] In situ hybridizations were performed for approximately 80 genes (Figure 3 and Supplemental Table 2; supplementary material online at www.gastrojournal.org). All tested genes were expressed in the Apc^{min} adenomas. The staining patterns in crypts could be grouped in several categories.

The first category consisted of about 80% of the tested genes. These genes were expressed in the proliferative compartment of the crypts. Examples are given in Figure 3*A–D*.

The second category of TCF target genes comprised the Paneth cell maturation markers.[23,24] The tyrosine kinase receptor EphB3, present in our Wnt/TCF signature gene set, falls in this category.[26] Other Paneth cell–specific genes such as *MMP7* (Figure 3*E* and *F*) and defensin-6 are not represented in the signature because they were not expressed in the cell lines. They were, however, clearly up-regulated in the tumor samples in the current study.

A third category (17 genes; Supplementary Table 2) yielded staining in 1–5 crypt cells, typically located near the crypt bottom. The cells were distinct from the Paneth cells. The location was highly reminiscent of the position to which the elusive crypt stem cells have been mapped.[27,28] As an example, the expression of the mouse ortholog of *ASCL2* is given in Figure 3*G* and *H*. Of note, a previously published intestinal stem cell marker, *Musashi*,[29,30] in our hands (data not shown) would be best classified in our first category.

Discussion

The current study builds on previous cell line–based work from our laboratory.[8] Here, we provide a comprehensive identification of TCF4 targets in 2 different cell lines carrying 2 different dnTCF genes, and by performing differential gene-expression analysis on a genome-wide oligonucleotide array platform. Moreover, we relate these findings to expression profiles of a set of human adenomas and adenocarcinomas. The Wnt/TCF signature gene set defines the core program activated by TCF4 in intestinal epithelial cells. Because the CRC cell lines used in this study arrest in the G1 phase of the cell cycle on inhibition of the Wnt cascade, this program is essential for the proliferative capacity of CRC cell lines in culture. These observations can be extrapolated to human intestinal tumors, in that the shared TCF4 target gene program likely represents the primary driver behind the transformation behavior of these transformed lesions. Moreover, the individual genes within the signature represent promising targets for therapy of CRC because their expression consistently is activated as the direct result of oncogenic Wnt pathway mutations, although many target genes will be involved causally in the transformed behavior of the neoplastic cells.

The TCF4 target gene program as activated in colorectal neoplasia consists of at least 3 distinct modules, as revealed by studies of physiologic gene expression in murine small intestinal crypts. Genes within one of the modules are expressed by the rapidly dividing crypt progenitors of the transit-amplifying compartment, and are likely the driving force behind the proliferative activity of intestinal neoplasia and cell lines derived thereof. Another module is correlated with maturation of postmitotic Paneth cells, as previously described.[24] We believe that the expression of the Paneth cell module in intestinal neoplasia is fortuitous and does not contribute to malignant transformation.

The discovery of a module expressed at the stem cell position was unexpected. Genes within this module may serve as starting points to study intestinal stem cell biology. A role for Wnt signaling in the biology of the transient-amplifying compartment of crypts has been firmly established previously.[5–7] A similar role in the biology of the intestinal stem cell has so far remained speculative.[31] The TCF target genes in the stem cell module may solidly link physiologic Wnt signaling to intestinal stem cell biology.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1053/j.gastro.2006.08.039.

References

1. Kinzler KW, Vogelstein B. Lessons from hereditary colorectal cancer. Cell 1996;87:159–170.
2. Nusse R. Wnt signaling in disease and in development. Cell Res 2005;15:28–32.
3. Korinek V, Barker N, Morin PJ, van Wichen D, de Weger R, Kinzler KW, Vogelstein B, Clevers H. Constitutive transcriptional activation by a beta-catenin-Tcf complex in APC-/- colon carcinoma. Science 1997;275:1784–1787.
4. Morin PJ, Sparks AB, Korinek V, Barker N, Clevers H, Vogelstein B, Kinzler KW. Activation of beta-catenin-Tcf signaling in colon cancer by mutations in beta-catenin or APC. Science 1997;275:1787–1790.
5. Korinek V, Barker N, Moerer P, van Donselaar E, Huls G, Peters PJ, Clevers H. Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4. Nat Genet 1998;19:379–383.
6. Kuhnert F, Davis CR, Wang HT, Chu P, Lee M, Yuan J, Nusse R, Kuo CJ. Essential requirement for Wnt signaling in proliferation of adult small intestine and colon revealed by adenoviral expression of Dickkopf-1. Proc Natl Acad Sci U S A 2004;101:266–271.
7. Pinto D, Gregorieff A, Begthel H, Clevers H. Canonical Wnt signals are essential for homeostasis of the intestinal epithelium. Genes Dev 2003;17:1709–1713.
8. van de Wetering M, Sancho E, Verweij C, de Lau W, Oving I, Hurlstone A, van der Horn K, Batlle E, Coudreuse D, Haramis AP, Tjon-Pon-Fong M, Moerer P, van den Born M, Soete G, Pals S, Eilers M, Medema R, Clevers H. The beta-catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. Cell 2002;111:241–250.
9. van de Wetering M, Barker N, Harkes IC, van der Heyden M, Dijk NJ, Hollestelle A, Klijn JG, Clevers H, Schutte M. Mutant E-cadherin breast cancer cells do not display constitutive Wnt signaling. Cancer Res 2001;61:278–284.
10. di Pietro M, Bellver JS, Menigatti M, Bannwart F, Schnider A, Russell A, Truninger K, Jiricny J, Marra G. Defective DNA mismatch repair determines a characteristic transcriptional profile in proximal colon cancers. Gastroenterology 2005;129:1047–1059.
11. Gregorieff A, Pinto D, Begthel H, Destree O, Kielman M, Clevers H. Expression pattern of Wnt signaling components in the adult intestine. Gastroenterology 2005;129:626–638.
12. Molenaar M, van de Wetering M, Oosterwegel M, Peterson-Maduro J, Godsave S, Korinek V, Roose J, Destree O, Clevers H. XTcf-3 transcription factor mediates beta-catenin-induced axis formation in Xenopus embryos. Cell 1996;86:391–399.
13. Barker N, Huls G, Korinek V, Clevers H. Restricted high level expression of Tcf-4 protein in intestinal and mammary gland epithelium. Am J Pathol 1999;154:29–35.
14. Jho EH, Zhang T, Domon C, Joo CK, Freund JN, Costantini F. Wnt/beta-catenin/Tcf signaling induces the transcription of Axin2, a negative regulator of the signaling pathway. Mol Cell Biol 2002;22:1172–1183.
15. Lustig B, Jerchow B, Sachs M, Weiler S, Pietsch T, Karsten U, van de Wetering M, Clevers H, Schlag PM, Birchmeier W, Behrens J. Negative feedback loop of Wnt signaling through upregulation of conductin/axin2 in colorectal and liver tumors. Mol Cell Biol 2002;22:1184–1193.
16. Yan D, Wiesmann M, Rohan M, Chan V, Jefferson AB, Guo L, Sakamoto D, Caothien RH, Fuller JH, Reinhard C, Garcia PD, Randazzo FM, Escobedo J, Fantl WJ, Williams LT. Elevated expression of axin2 and hnkd mRNA provides evidence that Wnt/beta-catenin signaling is activated in human colon tumors. Proc Natl Acad Sci U S A 2001;98:14973–14978.
17. He TC, Sparks AB, Rago C, Hermeking H, Zawel L, da Costa LT, Morin PJ, Vogelstein B, Kinzler KW. Identification of c-MYC as a target of the APC pathway. Science 1998;281:1509–1512.
18. Jubb AM, Chalasani S, Frantz GD, Smits R, Grabsch HI, Kavi V, Maughan NJ, Hillan KJ, Quirke P, Koeppen H. Achaete-scute like 2 (ascl2) is a target of Wnt signalling and is upregulated in intestinal neoplasia. Oncogene 2006;25:3445–3457.
19. Yamamoto Y, Sakamoto M, Fujii G, Tsuiji H, Kenetaka K, Asaka M, Hirohashi S. Overexpression of orphan G-protein-coupled receptor, Gpr49, in human hepatocellular carcinomas with beta-catenin mutations. Hepatology 2003;37:528–533.
20. Kenny PA, Enver T, Ashworth A. Receptor and secreted targets of Wnt-1/beta-catenin signalling in mouse mammary epithelial cells. BMC Cancer 2005;5:3.
21. Kim JS, Crooks H, Dracheva T, Nishanian TG, Singh B, Jen J, Waldman T. Oncogenic beta-catenin is required for bone morphogenetic protein 4 expression in human cancer cells. Cancer Res 2002;62:2744–2748.
22. Reichling T, Goss KH, Carson DJ, Holdcraft RW, Ley-Ebert C, Witte D, Aronow BJ, Groden J. Transcriptional profiles of intestinal tumors in Apc(Min) mice are unique from those of embryonic intestine and identify novel gene targets dysregulated in human colorectal tumors. Cancer Res 2005;65:166–176.
23. Andreu P, Colnot S, Godard C, Gad S, Chafey P, Niwa-Kawakita M, Laurent-Puig P, Kahn A, Robine S, Perret C, Romagnolo B. Crypt-restricted proliferation and commitment to the Paneth cell lineage following Apc loss in the mouse intestine. Development 2005;132:1443–1451.
24. van Es JH, Jay P, Gregorieff A, van Gijn ME, Jonkheer S, Hatzis P, Thiele A, van den Born M, Begthel H, Brabletz T, Taketo MM, Clevers H. Wnt signalling induces maturation of Paneth cells in intestinal crypts. Nat Cell Biol 2005;7:381–386.
25. Su LK, Kinzler KW, Vogelstein B, Preisinger AC, Moser AR, Luongo C, Gould KA, Dove WF. Multiple intestinal neoplasia caused by a mutation in the murine homolog of the APC gene. Science 1992;256:668–670.
26. Batlle E, Henderson JT, Beghtel H, van den Born MM, Sancho E, Huls G, Meeldijk J, Robertson J, van de Wetering M, Pawson T, Clevers H. Beta-catenin and TCF mediate cell positioning in the intestinal epithelium by controlling the expression of EphB/ephrinB. Cell 2002;111:251–263.
27. Bjerknes M, Cheng H. Clonal analysis of mouse intestinal epithelial progenitors. Gastroenterology 1999;116:7–14.
28. Potten CS. Extreme sensitivity of some intestinal crypt cells to X and gamma irradiation. Nature 1977;269:518–521.
29. Kayahara T, Sawada M, Takaishi S, Fukui H, Seno H, Fukuzawa H, Suzuki K, Hiai H, Kageyama R, Okano H, Chiba T. Candidate markers for stem and early progenitor cells, Musashi-1 and Hes1, are expressed in crypt base columnar cells of mouse small intestine. FEBS Lett 2003;535:131–135.
30. Potten CS, Booth C, Tudor GL, Booth D, Brady G, Hurley P, Ashton G, Clarke R, Sakakibara S, Okano H. Identification of a putative intestinal stem cell and early lineage marker; musashi-1. Differentiation 2003;71:28–41.
31. Reya T, Clevers H. Wnt signalling in stem cells and cancer. Nature 2005;434:843–850.

Received June 6, 2006. Accepted July 19, 2006.

Address requests for reprints to: Hans Clevers, Hubrecht Institute, Netherlands Institute for Developmental Biology, Uppsalalaan 8, 3584CT, Utrecht, The Netherlands. e-mail: clevers@niob.knaw.nl; fax: (31) 30-2121-801.

L.G.V.D.F. and J.S.-B. contributed equally to this article.

Supplemental Data Table 1: β-catenin/TCF target genes as selected with the procedure described in Figure 2

Target genes found increased both in adenomas and in carcinomas

	Fold differences *				
Affymetrix ID	Adenomas	CRC	Symbol	GenBank	Name
212942_s_at	54.8	38.8	KIAA1199	AB033025	KIAA1199 protein
227475_at	24.4	32.9	FOXQ1	AI676059	forkhead box Q1
223509_at	16.1	19.0	CLDN2	AF177340	claudin 2
1554576_a_at	5.4	14.4	ETV4	BC007242	ets variant gene 4 (E1A enhancer binding protein, E1AF)
225842_at	5.2	15.2	PHLDA1	AK026181	pleckstrin homology-like domain, family A, member 1
226446_at	4.6	7.9	HES6	AW249678	hairy and enhancer of split 6 (Drosophila)
212014_x_at	4.4	4.0	CD44	AI493245	CD44 antigen (homing function and Indian blood group system)
201195_s_at	4.3	12.3	SLC7A5	AB018009	solute carrier family 7 (cationic amino acid transporter, y+ system), member 5
201563_at	4.0	2.7	SORD	L29008	sorbitol dehydrogenase
218704_at	3.8	4.5	FLJ20315	NM_017763	hypothetical protein FLJ20315
202613_at	3.8	5.6	CTPS	NM_001905	CTP synthase
201340_s_at	3.6	2.5	ENC1	AF010314	ectodermal-neural cortex (with BTB-like domain)
41037_at	3.5	5.8	TEAD4	U63824	TEA domain family member 4
203510_at	3.4	3.7	MET	BG170541	met proto-oncogene (hepatocyte growth factor receptor)
202431_s_at	3.2	4.4	MYC	NM_002467	v-myc myelocytomatosis viral oncogene homolog (avian)
216913_s_at	3.1	5.5	KIAA0690	AK021460	KIAA0690 protein
224468_s_at	3.0	3.0	MGC13170	BC006151	multidrug resistance-related protein
225387_at	3.0	6.8	TM4SF9	AA059445	transmembrane 4 superfamily member 9
204695_at	2.9	4.5	CDC25A	AI343459	cell division cycle 25A
1555758_a_at	2.8	2.5	CDKN3	AF213040	cyclin-dependent kinase inhibitor 3
203612_at	2.8	3.1	BYSL	NM_004053	bystin-like
200832_s_at	2.7	4.8	SCD	AB032261	stearoyl-CoA desaturase (delta-9-desaturase)
202936_s_at	2.7	3.1	SOX9	NM_000346	SRY (sex determining region Y)-box 9
224467_s_at	2.7	2.7	MGC13096	BC006146	hypothetical protein MGC13096
201416_at	2.7	4.0	SOX4	BG528420	SRY (sex determining region Y)-box 4
201420_s_at	2.7	2.5	WDR77	BF975273	WD repeat domain 77
210463_x_at	2.7	4.7	TRMT1	BC002492	TRM1 tRNA methyltransferase 1 homolog (S. cerevisiae)
223018_at	2.6	2.6	NOB1P	BC000050	likely ortholog of mouse nin one binding protein
206102_at	2.6	3.6	PSF1	NM_021067	DNA replication complex GINS protein PSF1
203867_s_at	2.5	3.3	NLE1	NM_018096	notchless homolog 1 (Drosophila)
213226_at	2.5	2.4	PMSCL1	AI346350	polymyositis/scleroderma autoantigen 1, 75kDa
220198_s_at	2.4	4.3	EIF5A2	NM_020390	eukaryotic translation initiation factor 5A2
203313_s_at	2.4	2.0	TGIF	NM_003244	TGFB-induced factor (TALE family homeobox)
222500_at	2.4	2.2	PPIL1	BC003048	peptidylprolyl isomerase (cyclophilin)-like 1
201329_s_at	2.4	2.0	ETS2	NM_005239	v-ets erythroblastosis virus E26 oncogene homolog 2 (avian)
225295_at	2.4	3.9	SLC39A10	AB033091	solute carrier family 39 (zinc transporter), member 10
204133_at	2.4	2.9	RNU3IP2	NM_004704	RNA, U3 small nucleolar interacting protein 2
230656_s_at	2.3	2.2	CIRH1A	AL578336	cirrhosis, autosomal recessive 1A (cirhin)
225307_at	2.3	3.3	ZNF511	AL583632	zinc finger protein 511
208152_s_at	2.3	3.4	DDX21	NM_004728	DEAD (Asp-Glu-Ala-Asp) box polypeptide 21
223151_at	2.3	3.1	DCUN1D5	BC004169	DCN1, defective in cullin neddylation 1, domain containing 5 (S. cerevisiae)
218507_at	2.3	6.2	HIG2	NM_013332	hypoxia-inducible protein 2
204175_at	2.3	3.2	ZNF593	NM_015871	zinc finger protein 593
220688_s_at	2.3	2.3	C1orf33	NM_016183	chromosome 1 open reading frame 33
201478_s_at	2.3	3.2	DKC1	U59151	dyskeratosis congenita 1, dyskerin
200790_at	2.3	2.5	ODC1	NM_002539	ornithine decarboxylase 1
202246_s_at	2.3	2.7	CDK4	NM_000075	cyclin-dependent kinase 4
228217_s_at	2.3	2.5	LOC389362	BF973374	hypothetical LOC389362
221649_s_at	2.2	3.7	PPAN	BC000535	peter pan homolog (Drosophila)
212279_at	2.2	2.5	MAC30	BE779865	hypothetical protein MAC30
225799_at	2.2	6.1	MGC4677	BF209337	hypothetical protein MGC4677, hypothetical LOC541471
203023_at	2.2	2.1	HSPC111	NM_016391	hypothetical protein HSPC111
209464_at	2.2	2.3	AURKB	AB011446	aurora kinase B
226349_at	2.2	2.0	MGC40397	BE264828	Hypothetical protein MGC40397
218529_at	2.2	2.4	CD320	NM_016579	CD320 antigen
222760_at	2.1	2.4	ZNF703	BG290193	zinc finger protein 703
218512_at	2.1	2.3	WDR12	NM_018256	WD repeat domain 12
217850_at	2.1	2.5	NS	NM_014366	nucleostemin
209832_s_at	2.1	2.7	CDT1	AF321125	DNA replication factor
203622_s_at	2.1	2.6	LOC56902	NM_020143	putatative 28 kDa protein
221514_at	2.1	2.6	SDCCAG16	BC001149	serologically defined colon cancer antigen 16
218889_at	2.1	2.5	NOC3L	NM_022451	nucleolar complex associated 3 homolog (S. cerevisiae)
218590_at	2.1	2.8	PEO1	NM_021830	progressive external ophthalmoplegia 1

1

	Fold differences *				
Affymetrix ID	Adenomas	CRC	Symbol	GenBank	Name
202715_at	2.0	3.0	CAD	NM_004341	carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase
212160_at	2.0	2.8	XPOT	AI984005	exportin, tRNA (nuclear export receptor for tRNAs)
223413_s_at	2.0	2.5	FLJ20425	AW958593	hypothetical protein FLJ20425
207515_s_at	1.9	1.9	POLR1C	NM_004875	polymerase (RNA) I polypeptide C, 30kDa
217884_at	1.9	2.4	FLJ10774	NM_024662	N-acetyltransferase-like protein
224428_s_at	1.9	2.3	CDCA7	AY029179	cell division cycle associated 7
221712_s_at	1.9	2.1	WDR74	BC006351	WD repeat domain 74
213342_at	1.9	2.4	YAP1	AI745185	Yes-associated protein 1, 65kDa
203114_at	1.9	2.3	MTVR1	NM_006396	Mouse Mammary Turmor Virus Receptor homolog 1
201516_at	1.8	2.7	SRM	NM_003132	spermidine synthase
218708_at	1.8	2.5	NXT1	NM_013248	NTF2-like export factor 1
225699_at	1.8	5.1	LOC285958	AI937446	hypothetical protein LOC285958. ACA9 snoRNA gene
227103_s_at	1.8	2.0	MGC2408	BE646208	hypothetical protein MGC2408
224331_s_at	1.8	1.6	MRPL36	AB049654	mitochondrial ribosomal protein L36
204977_at	1.8	2.2	DDX10	NM_004398	DEAD (Asp-Glu-Ala-Asp) box polypeptide 10
210008_s_at	1.8	2.7	MRPS12	AA513737	mitochondrial ribosomal protein S12
203150_at	1.8	1.7	RAB9P40	NM_005833	Rab9 effector p40
201453_x_at	1.8	2.3	RHEB	NM_005614	Ras homolog enriched in brain
218882_s_at	1.8	1.8	WDR3	NM_006784	WD repeat domain 3
210541_s_at	1.8	1.8	RFP	AF230394	ret finger protein
223403_s_at	1.7	1.8	POLR1B	BC004882	polymerase (RNA) I polypeptide B, 128kDa
224634_at	1.7	2.2	FLJ20249	AI911518	hypothetical protein FLJ20249
223331_s_at	1.7	1.6	DDX20	AF106019	DEAD (Asp-Glu-Ala-Asp) box polypeptide 20
224714_at	1.7	1.7	MKI67IP	AL542544	MKI67 (FHA domain) interacting nucleolar phosphoprotein
202159_at	1.7	1.7	FARSLA	NM_004461	phenylalanine-tRNA synthetase-like, alpha subunit
1554464_a_at	1.7	1.7	CRTAP	BC008745	cartilage associated protein
37462_i_at	1.7	2.5	SF3A2	L21990	splicing factor 3a, subunit 2, 66kDa
225712_at	1.7	2.5	GEMIN5	AW024563	gem (nuclear organelle) associated protein 5
218258_at	1.7	1.9	POLR1D	NM_015972	polymerase (RNA) I polypeptide D, 16kDa
204027_s_at	1.7	1.9	METTL1	NM_005371	methyltransferase-like 1
212846_at	1.7	2.3	KIAA0179	AA811192	KIAA0179 protein
200054_at	1.6	1.8	ZNF259	NM_003904	zinc finger protein 259
218695_at	1.6	2.0	RRP41	NM_019037	exosome complex exonuclease RRP41
1554678_s_at	1.6	2.9	HNRPDL	AB066484	heterogeneous nuclear ribonucleoprotein D-like
210672_s_at	1.6	2.8	C16orf35	BC004185	chromosome 16 open reading frame 35
203119_at	1.6	2.3	MGC2574	NM_024098	hypothetical protein MGC2574
217754_at	1.6	2.2	DDX56	NM_019082	DEAD (Asp-Glu-Ala-Asp) box polypeptide 56
207891_s_at	1.6	1.9	TREX2	NM_017518	three prime repair exonuclease 2
219110_at	1.6	1.4	NOLA1	NM_018983	nucleolar protein family A, member 1 (H/ACA small nucleolar RNPs)
33307_at	1.6	1.8	CGI-96	AL022316	CGI-96 protein
212411_at	1.6	1.5	IMP4	BE747342	IMP4, U3 small nucleolar ribonucleoprotein, homolog (yeast)
218398_at	1.6	1.7	MRPS30	NM_016640	mitochondrial ribosomal protein S30
222703_s_at	1.6	1.9	FLJ23476	BE464161	ischemia/reperfusion inducible protein
218594_at	1.6	1.9	HEATR1	NM_018072	HEAT repeat containing 1
204717_s_at	1.6	2.0	SLC29A2	AF034102	solute carrier family 29 (nucleoside transporters), member 2
219131_at	1.6	1.7	TERE1	NM_013319	transitional epithelia response protein
225196_s_at	1.6	1.8	MRPS26	AF308301	mitochondrial ribosomal protein S26
211576_s_at	1.5	2.2	SLC19A1	BC003068	solute carrier family 19 (folate transporter), member 1
227042_at	1.5	1.7	LOC150223	BE218514	hypothetical protein LOC150223
225402_at	1.5	2.3	C20orf64	BG339450	chromosome 20 open reading frame 64
219575_s_at	1.5	1.6	COG8	NM_022341	component of oligomeric golgi complex 8
220147_s_at	1.5	1.6	C12orf14	NM_021238	chromosome 12 open reading frame 14
219098_at	1.5	1.8	MYBBP1A	NM_014520	MYB binding protein (P160) 1a
212422_at	1.4	1.6	PDCD11	AL547263	programmed cell death 11
214427_at	1.4	2.0	NOL1	NM_006170	nucleolar protein 1, 120kDa
200874_s_at	1.4	1.5	NOL5A	BE796327	nucleolar protein 5A (56kDa with KKE/D repeat)
200924_s_at	1.4	2.3	SLC3A2	NM_002394	solute carrier family 3 (activators of dibasic and neutral amino acid transport), 2
201270_x_at	1.4	1.9	KIAA1068	NM_015332	KIAA1068 protein

2

Target genes found increased only in adenomas

Affy ID	Fold difference*	Symbol	GenBank	Name
207607_at	17.3	ASCL2	NM_005170	achaete-scute complex-like 2 (Drosophila)
224176_s_at	6.6	AXIN2	AF205888	axin 2 (conductin, axil)
213880_at	5.8	GPR49	AL524520	G protein-coupled receptor 49
206286_s_at	4.7	TDGF1	NM_003212	teratocarcinoma-derived growth factor 1
229576_s_at	4.6	TBX3	N29712	T-box 3 (ulnar mammary syndrome)
1438_at	3.3	EPHB3	X75208	EPH receptor B3
209589_s_at	3.1	EPHB2	AF025304	EPH receptor B2
226360_at	2.6	ZNRF3	AK022809	zinc and ring finger 3
223253_at	2.5	UCC1	BC000686	upregulated in colorectal cancer gene 1
222829_s_at	2.5	IL20RA	BE219979	interleukin 20 receptor, alpha
225202_at	2.5	RHOBTB3	BE620739	Rho-related BTB domain containing 3
206628_at	2.5	SLC5A1	NM_000343	solute carrier family 5 (sodium/glucose cotransporter), member 1
227607_at	2.2	AMSH-LP	AI638611	associated molecule with the SH3 domain of STAM (AMSH) like protein
205527_s_at	2.2	GEMIN4	NM_015487	gem (nuclear organelle) associated protein 4
202831_at	2.1	GPX2	NM_002083	glutathione peroxidase 2 (gastrointestinal)
207826_s_at	2.0	ID3	NM_002167	inhibitor of DNA binding 3, dominant negative helix-loop-helix protein
217786_at	2.0	SKB1	NM_006109	SKB1 homolog (S. pombe)
222922_at	1.9	KCNE3	AF302494	potassium voltage-gated channel, Isk-related family, member 3
52285_f_at	1.8	C18orf9	AW002970	chromosome 18 open reading frame 9
219037_at	1.8	CGI-115	NM_016052	CGI-115 protein
219050_s_at	1.8	C11orf5	NM_014205	chromosome 11 open reading frame 5
225667_s_at	1.8	NSE1	AI601101	NSE1
209893_s_at	1.8	FUT4	M58596	fucosyltransferase 4 (alpha (1,3) fucosyltransferase, myeloid-specific)
222875_at	1.7	DHX33	AI720923	DEAH (Asp-Glu-Ala-His) box polypeptide 33
219961_s_at	1.7	C20orf19	NM_018474	chromosome 20 open reading frame 19
219539_at	1.7	GEMIN6	NM_024775	gem (nuclear organelle) associated protein 6
216268_s_at	1.7	JAG1	U77914	jagged 1 (Alagille syndrome)
1558292_s_at	1.7	PIGW	BF037819	phosphatidylinositol glycan, class W
204798_at	1.7	MYB	NM_005375	v-myb myeloblastosis viral oncogene homolog (avian)
210027_s_at	1.7	APEX1	M80261	APEX nuclease (multifunctional DNA repair enzyme) 1
228654_at	1.7	LOC139886	AU145277	hypothetical protein LOC139886
219031_s_at	1.6	CGI-37	NM_016101	comparative gene identification transcript 37
204405_x_at	1.6	HSA9761	NM_014473	putative dimethyladenosine transferase
218692_at	1.6	FLJ20366	NM_017786	hypothetical protein FLJ20366
200772_x_at	1.6	PTMA	BF686442	prothymosin, alpha (gene sequence 28)
213237_at	1.6	MGC16824	AI652058	esophageal cancer associated protein
223172_s_at	1.6	HSPC242	AF060924	hypothetical protein HSPC242
209494_s_at	1.5	ZNF278	AI807017	zinc finger protein 278
218957_s_at	1.5	FLJ11848	NM_025155	hypothetical protein FLJ11848
218140_x_at	1.5	SRPRB	NM_021203	signal recognition particle receptor, B subunit
228009_x_at	1.5	ZNRD1	NM_014596	zinc ribbon domain containing, 1
218605_at	1.5	TFB2M	NM_022366	transcription factor B2, mitochondrial
209382_at	1.5	RPC62	U93867	polymerase (RNA) III (DNA directed) (62kD)
212614_at	1.4	ARID5B	BG285011	AT rich interactive domain 5B (MRF1-like)
201054_at	1.4	HNRPA0	BE966599	heterogeneous nuclear ribonucleoprotein A0
225253_s_at	1.4	METTL2	AI632244	methyltransferase like 2
202138_x_at	1.4	JTV1	NM_006303	JTV1 gene
228614_at	1.4	LOC205251	AW182614	hypothetical protein LOC205251
218027_at	1.4	MRPL15	NM_014175	mitochondrial ribosomal protein L15
211778_s_at	1.4	ZNF339	BC006148	zinc finger protein 339
203931_s_at	1.3	MRPL12	NM_002949	mitochondrial ribosomal protein L12

3

Target genes found increased only in carcinomas

Affy ID	Fold difference*	Symbol	GenBank	Name
214974_x_at	58.6	CXCL5	AK026546	chemokine (C-X-C motif) ligand 5
223642_at	44.0	ZIC2	AF193855	Zic family member 2 (odd-paired homolog, Drosophila)
226473_at	11.6	CBX2	BE514414	chromobox homolog 2 (Pc class homolog, Drosophila)
235182_at	10.8	C20orf82	AI816793	chromosome 20 open reading frame 82
1564706_s_at	4.6	GLS2	AF110329	glutaminase 2 (liver, mitochondrial)
209406_at	4.5	BAG2	AF095192	BCL2-associated athanogene 2
227195_at	4.1	ZNF503	AA603467	zinc finger protein 503
65588_at	4.0	LOC388796	AA827892	hypothetical LOC388796
221558_s_at	3.6	LEF1	AF288571	lymphoid enhancer-binding factor 1
225857_s_at	3.1	FLJ12683	AA827892	RNA, U71A small nucleolar
221214_s_at	2.8	NELF	NM_015537	nasal embryonic luteinizing hormone-releasing hormone factor
213823_at	2.6	HOXA11	H94842	homeo box A11
226017_at	2.4	CKLFSF7	AI708432	chemokine-like factor super family 7
212687_at	2.3	LIMS1	AL110164	LIM and senescent cell antigen-like domains 1
209681_at	2.1	SLC19A2	AF153330	solute carrier family 19 (thiamine transporter), member 2
59705_at	2.1	SCLY	AA911739	selenocysteine lyase
217427_s_at	2.1	HIRA	X75296	HIR histone cell cycle regulation defective homolog A (S. cerevisiae)
218151_x_at	2.1	GPR172A	NM_024531	G protein-coupled receptor 172A
205003_at	2.0	DOCK4	NM_014705	dedicator of cytokinesis 4
212360_at	2.0	AMPD2	AI916249	adenosine monophosphate deaminase 2 (isoform L)
200768_s_at	1.9	MAT2A	BC001686	methionine adenosyltransferase II, alpha
218712_at	1.9	FLJ20508	NM_017850	hypothetical protein FLJ20508
204079_at	1.9	TPST2	NM_003595	tyrosylprotein sulfotransferase 2
223204_at	1.9	DKFZp434L142	AF260333	hypothetical protein DKFZp434L142
201490_s_at	1.9	PPIF	NM_005729	peptidylprolyl isomerase F (cyclophilin F)
224966_s_at	1.9	LOC56931	AI857915	hypothetical protein from EUROIMAGE 1967720
212766_s_at	1.8	FLJ12671	AW294587	hypothetical protein FLJ12671
223222_at	1.7	SLC25A19	BC001075	solute carrier family 25 (mitochondrial deoxynucleotide carrier), member 19
208916_at	1.7	SLC1A5	AF105230	solute carrier family 1 (neutral amino acid transporter), member 5
1555764_s_at	1.7	TIMM10	AF152354	translocase of inner mitochondrial membrane 10 homolog (yeast)
203782_s_at	1.7	POLRMT	NM_005035	polymerase (RNA) mitochondrial (DNA directed)
210213_s_at	1.6	ITGB4BP	AF022229	integrin beta 4 binding protein
225383_at	1.6	ZNF275	BF793625	zinc finger protein 275
214141_x_at	1.6	SFRS7	BF033354	splicing factor, arginine/serine-rich 7, 35kDa
212168_at	1.5	RBM12	AL514547	RNA binding motif protein 12
209825_s_at	1.4	UMPK	BC002906	uridine monophosphate kinase

* Fold differences in the transcript levels in adenomas or carcinomas *versus* normal mucosa. The non-parametric Mann-Whitney test was used with a false discovery rate of 0.05 and Bonferroni correction for the group comparison analysis.

IMP4	BE747342	IMP4, U3 small nucleolar ribonucleoprotein, homolog (yeast)	1
HEATR1	NM_018072	HEAT repeat containing 1	1
SLC19A1	BC003068	solute carrier family 19 (folate transporter), member 1	1
C20orf64	BG339450	chromosome 20 open reading frame 64	1
C12orf14	NM_021238	chromosome 12 open reading frame 14	1
PDCD11	AL547263	programmed cell death 11	1
NOL1	NM_006170	nucleolar protein 1, 120kDa	1
ASCL2	NM_005170	achaete-scute complex-like 2 (Drosophila)	3
TDGF1	NM_003212	teratocarcinoma-derived growth factor 1	3
TBX3	N29712	T-box 3 (ulnar mammary syndrome)	3
EPHB3	X75208	EPH receptor B3	2
ZNRF3	AK022809	zinc and ring finger 3	3
IL20RA	BE219979	interleukin 20 receptor, alpha	1
RHOBTB3	BE620739	Rho-related BTB domain containing 3	3
GEMIN4	NM_015487	gem (nuclear organelle) associated protein 4	3
ID3	NM_002167	inhibitor of DNA binding 3, dominant negative helix-loop-helix protein	1
SKB1	NM_006109	SKB1 homolog (S. pombe)	1
C11orf5	NM_014205	chromosome 11 open reading frame 5	1
FUT4	M58596	fucosyltransferase 4 (alpha (1,3) fucosyltransferase, myeloid-specific)	1
PIGW	BF037819	phosphatidylinositol glycan, class W	1
APEX1	M80261	APEX nuclease (multifunctional DNA repair enzyme) 1	3
MGC16824	AI652058	esophageal cancer associated protein	3
ZNF278	AI807017	zinc finger protein 278	3
SRPRB	NM_021203	signal recognition particle receptor, B subunit	1
ZNRD1	NM_014596	zinc ribbon domain containing, 1	1
TFB2M	NM_022366	transcription factor B2, mitochondrial	1
RPC62	U93867	polymerase (RNA) III (DNA directed) (62kD)	1
JTV1	NM_006303	JTV1 gene	1
ZNF339	BC006148	zinc finger protein 339	1

Wnt signature genes tested with *in situ* hybridizations were expressed in wt crypts and in adenomas of Apc^{min} mice. The staining patterns in crypts could be divided into three categories. Genes within module 1 are expressed by the rapidly dividing crypt progenitors of the transit amplifying compartment. Module 2 is correlated with maturation of post-mitotic Paneth cells. Module 3 consists out of the target genes expressed at the stem cell position in the crypts.

6

6.3. Appendix III

"Defective DNA Mismatch Repair Determines a Characteristic Transcriptional Profile in Proximal Colon Cancers" Massimiliano di Pietro*, Jacob Sabates Bellver*, Mirco Menigatti, Fridolin Bannwart, Annelies Schnider, Anna Russell, Kaspar Truninger, Josef Jiricny and Giancarlo Marra. Gastroenterology 2005;129:1047-59. *These authors contribute equally to this work

GASTROENTEROLOGY 2005;129:1047–1059

MICROARRAYS AND OTHER NEW TECHNOLOGIES

Defective DNA Mismatch Repair Determines a Characteristic Transcriptional Profile in Proximal Colon Cancers

MASSIMILIANO DI PIETRO,* JACOB SABATES BELLVER,* MIRCO MENIGATTI,* FRIDOLIN BANNWART,[†] ANNELIES SCHNIDER,[§] ANNA RUSSELL,[‖] KASPAR TRUNINGER,[¶] JOSEF JIRICNY,* and GIANCARLO MARRA*

*Institute of Molecular Cancer Research, University of Zurich, Zurich; [†]Institute of Pathology and [§]Department of Surgery, Triemli Hospital, Zurich; [‖]Research Group Human Genetics, Center for Biomedicine, University of Basel, Basel; and [¶]Division of Gastroenterology, Cantonal Hospital, Aarau, Switzerland

Background & Aims: **Colon cancers with defective DNA mismatch repair (MMR) have peculiar molecular, pathologic, and clinical features, including high-level microsatellite instability, conspicuous lymphocytic infiltration, preferential location in the proximal colon, and better prognosis. Our aim was to characterize the transcriptional profile of this colon cancer subset.** ***Methods:*** **An oligonucleotide microarray containing 12,625 probes was used to evaluate gene expression in 25 proximal colon cancers, 10 samples of normal colon mucosa, and 14 colon cancer cell lines. Transcriptional profiles of MMR-deficient cancers and cell lines were compared with those of their MMR-proficient counterparts.** ***Results:*** **Unsupervised analysis of microarray data showed that MMR status exerts a predominant influence on the gene expression profile of proximal colon cancers. Hierarchical clustering divided the cancers into 2 groups corresponding almost perfectly with their MMR status. Supervised analysis identified numerous gene expression changes that represent a genetic signature of MMR-deficient colon cancers. Changes in genes involved in apoptosis and the immune response were consistent with the better prognosis of MMR-deficient cancers. In MMR-deficient cancers and cell lines, *4-1BBL*, a crucial gene in the anti-tumor immune response, was, respectively, 2.4 and 6.0 times more expressed than in their MMR-proficient counterparts. This difference was confirmed by quantitative reverse-transcription polymerase chain reaction and flow cytometric assessment of 4-1BBL protein expression in colon cancer cell lines. Our analysis also showed novel possible gene targets of microsatellite instability.** ***Conclusions:*** **MMR inactivation produces distinct changes in the cellular messenger RNA pool, which is consistent with a unique tumorigenesis pathway.**

Mismatch repair (MMR)-deficient colorectal cancers are associated with inherited[1] and somatic[2] alterations in the *MSH2*, *MSH6*, *MLH1*, and *PMS2* genes, which are involved in the repair of DNA base-base mispairs and strand misalignments that occur during DNA replication.[3,4] MMR-deficient cells are phenotypically characterized by microsatellite instability (MSI), widespread genetic instability involving mononucleotide and dinucleotide repeats.[5] Its tumorigenic relevance is supported by evidence that malignant transformation may depend on frameshift mutations within microsatellites in the coding region of tumor suppressor genes.[6,7] Compared with their MMR-proficient counterparts, MMR-deficient colon cancers are more frequent in the proximal colon and more likely to be near diploid, poorly differentiated, and mucinous. They also have lower rates of lymph node metastasis and a better overall prognosis, which is believed to depend in part on their seemingly enhanced immunogenicity reflected by conspicuous intraepithelial T-lymphocyte infiltrations and nodular B-cell aggregation.[8–10] It is now accepted that MMR-deficient colon cancers arise through a distinct transformation process. Our objective was to identify the transcriptional features that distinguish these cancers from those in which the MMR system is fully functional.

Materials and Methods

We analyzed 25 primary proximal colon cancers (12 MMR deficient and 13 MMR proficient) (Table 1), 10 specimens of normal proximal colon mucosa (from 5 of the 12 patients with MMR-deficient colon cancers and 5 of the 13 patients with MMR-proficient colon cancers), and 14 colon cancer epithelial cell lines (7 MMR deficient and 7 MMR proficient) (Table 2). The MMR status of the cancers was determined by immunohistochemical assessment of *MSH2*,

Abbreviations used in this paper: MMR, mismatch repair; MSI, microsatellite instability; UTR, untranslated region.

0016-5085/05/$30.00
doi:10.1053/j.gastro.2005.06.028

Table 1. Clinical, Pathologic, and Molecular Characteristics of MMR(+) and MMR(−) Colon Cancer Cases

Tumor no.[a]	Age (y)	Sex	Site[b]	Stage[c]	Grade[c]	Lymphocyte infiltration[d]	Mucin (amount)[e]	Mucin (type)[f]	IHC[g]	BAT26	*MLH1* promoter[h]
MMR(+) 1	81	F	C	T2 N0	G2	+	(+)	n	Normal	Stable	
MMR(+) 2	90	F	C	T3 N0	G2	−	(+)	n	Normal	Stable	
MMR(+) 3	55	M	A	T3 N0	G3	−	−	—	Normal	Stable	
MMR(+) 4	80	F	A	T3 N1	G2	−	+	n	Normal	Stable	
MMR(+) 5	83	F	A	T4 N2	G3	+	+	n	Normal	Stable	
MMR(+) 6	78	M	A	T3 N0	G2	−	(+)	n	Normal	Stable	
MMR(+) 7	87	F	A	T3 N0	G2	−	+	n	Normal	Stable	
MMR(+) 8	73	M	A	T3 N1	G2	−	+	n	Normal	Stable	
MMR(+) 9	69	M	C	T3 N0	G2	+	+	n	Normal	Stable	
MMR(+) 10	77	M	A	T3 N0	G2	+	(+)	n	Normal	Stable	
MMR(+) 11	86	M	A	T2 N0	G2	−	+	n	Normal	Stable	
MMR(+) 12	83	F	C	T3 N0	G2	−	+	n	Normal	Stable	
MMR(+) 13	83	F	A	T3 N0	G2	−	(+)	n	Normal	Stable	
MMR(−) 1	81	F	A	T3 N0	G2	++	−	—	Lack of MLH1	Unstable	H
MMR(−) 2	75	F	A	T3 N0	G2	+	++	a	Lack of MLH1	Unstable	H
MMR(−) 3	54	M	A	T3 N0	G3	+	++	n	Lack of MLH1	Unstable	H
MMR(−) 4	80	F	A	T3 N0	G2	++	++	a	Lack of MLH1	Unstable	H
MMR(−) 5	90	M	A	T4 N1	G3	++	+	n	Lack of MLH1	Unstable	H
MMR(−) 6	81	M	HF	T4 N0	G3	+++	−	—	Lack of MLH1	Unstable	H
MMR(−) 7	77	F	C	T3 N0	G2	+++	++	a	Lack of MLH1	Unstable	H
MMR(−) 8	88	M	A	T3 N0	G3	−[i]	+	a	Lack of MLH1	Unstable	H
MMR(−) 9	85	F	A	T4 N2	G3	+	++	a	Lack of MLH1	Unstable	H
MMR(−) 10	77	M	HF	T3 N0	G2	++	++	a	Lack of MLH1	Unstable	H
MMR(−) 11	74	F	C	T3 N0	G2	++	+	n	Lack of MSH2	Unstable	Un
MMR(−) 12	21	M	A	T3 N0	G2	++	++	a	Lack of MSH2	Unstable	Un

MMR(+), MMR proficient; MMR(−), deficient; C, cecum; A, ascending; HF, hepatic flexure; H, hypermethylated; Un, unmethylated.

[a]All cancers were apparently sporadic, with the exception of MMR(−) 12, which was removed from a member of a hereditary nonpolyposis colorectal cancer–like family (ie, one satisfying the Bethesda Guidelines [J Natl Cancer Inst 2004:96:261] but not the Amsterdam criteria [Gastroenterology 1999:116:1453].

[b]All cancers were from the proximal colon.

[c]TNM and grade classification according to Sobin LH, Wittekind C. TNM classification of malignant tumours. 6th ed. New York, NY: Wiley-Liss, 2002.

[d]Semiquantitative evaluation of intraepithelial lymphocytic infiltration. −, <5 lymphocytes per high-power field; +, <10 lymphocytes per high-power field; ++, 10–20 lymphocytes per high-power field; +++, >20 lymphocytes per high-power field.

[e]Semiquantitative evaluation of mucin content; −, no detectable mucin; (+), low but detectable levels of intratubular mucin: +, intratubular mucin present; ++, intratubular mucin and mucin lakes.

[f]Dominant type of mucin; n, neutral; a, acid; as detected by periodic acid–Schiff and Alcian blue staining, respectively.

[g]Immunohistochemistry for the MMR proteins MSH2, MSH6, MLH1, and PMS2.

[h]Methylation-specific polymerase chain reaction analysis of the *MLH1* promoter was performed only in the 12 MMR(−) cancers.

[i]Signet ring cell carcinoma.

MSH6, *MLH1*, and *PMS2* expression and by evaluation of MSI at BAT-26 and *hMLH1* promoter methylation, as described elsewhere.[4] The MMR status of the cell lines, which has been documented in previous reports, was confirmed by Western blot of the MMR proteins.[11]

All tissues were prospectively collected at a single hospital in Zurich, Switzerland, with full institutional review board approval. Immediately after resection, cancers were trimmed (ulcerated and necrotic tissues, adenomatous tissue margins), and the resulting specimen, representing the invasive region of the cancer, was cut into smaller pieces and stored in RNAlater (Ambion, Huntingdon, UK) for microarray analysis. The tissue adjacent to this section was fixed in buffered formalin, embedded in paraffin, and sectioned for histologic and molecular evaluation of tumor characteristics (Table 1). This area was also used to evaluate the percentage of cancerous epithelial cells in each tumor (range, 50%–85%).

Fifty milligrams of the fragments stored in RNAlater were homogenized, total RNA was extracted (Absolutely RNA kit; Stratagene, Amsterdam, The Netherlands), and RNA integrity was verified by capillary gel electrophoresis. The same procedure (excluding homogenization) was used to extract total RNA from 5×10^6 cultured cells of each line. For each tissue or cell line specimen, 15 μg of complementary RNA, synthesized and labeled as previously described,[12] were analyzed with the Affymetrix U95Av2 array (Affymetrix, High Wycombe, UK), which contains in situ synthesized oligonucleotides (12,625 probes) representing approximately 10,000 unique full-length genes. Gene expression data were analyzed with unsupervised and supervised hierarchical clustering analyses and principal component analysis (see supplemental material online at www.gastrojournal.org and http://www.imcr.unizh.ch/research/researchdetail.php?StaffId=11). GeneSpring software (Silicon Ge-

Table 2. Characteristics of the 14 Colon Cancer Cell Lines Studied

Cell lines	Western blot (steady-state protein levels)				*MLH1* promoter methylation[a]
	MSH2	MSH6	MLH1	PMS2	
MMR+					
SW480	Normal	Normal	Normal	Normal	Unmethylated
SW620	Normal	Normal	Normal	Normal	Unmethylated
SW837	Normal	Normal	Normal	Normal	Unmethylated
COLO741	Normal	Normal	Normal	Normal	Unmethylated
Caco2	Normal	Normal	Normal	Normal	Unmethylated
HT29	Normal	Normal	Normal	Normal	Unmethylated
CX-1	Normal	Normal	Normal	Normal	Unmethylated
MMR−					
CO115	Low	Low	**Absent**	Very weak	Hypermethylated
SW48	Normal	Normal	**Absent**	Very weak	Hypermethylated
LS174T	Low	Low	**Very weak**	Very weak	Unmethylated[b]
LS180	Low	Low	**Very weak**	Very weak	Unmethylated[b]
LS411	Normal	Normal	**Absent**	Very weak	Hypermethylated
HCT116	Normal	Normal	**Absent**	Very weak	Unmethylated[c]
GP5d	**Absent**	Very weak	Normal	Normal	Unmethylated

NOTE. COLO741 and CX1 cells were kindly provided by Ian Tomlinson (London Research Institute, England) and GP5d cells by Donna E. Davies (CRC Wessex Medical Oncology Unit, Southampton, England). The other cell lines were from the cell line repository of the Zurich Cancer Network. The primary alteration of MMR protein expression is reported in bold. MLH1 and MSH2 alterations lead to proteolytic degradation of PMS2 and MSH6, respectively. The nonexpression of MSH2 in GP5d cells is a novel finding; all other alterations are consistent with current literature data.
[a]Data in this column refer to methylation of the *MLH1* promoter region that is critical for *MLH1* gene silencing, that is, bases −289 to −202 upstream from the transcription start site (GeneBank no. U83845).[4]
[b]In these cell lines, the absence of methylation in the previously cited region was accompanied by methylation of another region (−716 to −563) of the promoter, which is considered to be less closely associated with gene silencing.[4]
[c]*MLH1* is mutated in HCT116.

netics, Redwood City, CA) was used for statistical analysis of microarray data. For supervised statistical group comparisons, the Mann–Whitney test was used with a false discovery rate of 0.05 and Benjamini–Hochberg multiple testing correction; for comparison of normal mucosal tissues and cancers, which yielded larger intergroup differences, the more stringent Bonferroni correction was used. Other data were analyzed with Student *t* test, χ^2 test, and Fisher exact test.

Microarray findings were confirmed by one-step real-time reverse-transcription polymerase chain reaction (as described previously[12]), and 4-1BBL protein expression in colon cancer cell lines was evaluated with flow cytometry (see supplemental material online at www.gastrojournal.org).

Results

Fifty-one proximal colon cancers were collected in 2001 and 2002. All 12 (23.5%) satisfying our criteria for MMR deficiency (see Materials and Methods) underwent microarray analysis (Table 1), along with 13 of 39 MMR-proficient cancers, the closest matches for the MMR-deficient tumors (in terms of patient age/sex, tumor stage/grade, percentage of neoplastic epithelial cells in the specimen).

Intraepithelial lymphocyte infiltration and mucin positivity were more frequent among MMR-deficient cancers. *MLH1* gene silencing accounted for the MMR deficiency in 10 of 12 MMR-deficient cancers (all sporadic). The mean *MLH1* transcript level in the MMR-proficient cancers was 2.2 times higher than that of the MLH1-deficient subset of MMR-deficient cancers ($t = 6.34$; $P < .0001$).

Unsupervised hierarchical clustering effectively segregated the normal mucosa and cancer specimens (Figure 1). Supervised statistical group comparison analysis then identified 182 genes with markedly different transcript levels in these 2 groups (fold differences of ≥4.0), suggesting their potential relevance to colon carcinogenesis in general (Supplementary Table 1, see supplemental material online at www.gastrojournal.org). The numerous genes in this set whose association with colon cancer has not been previously documented are the subjects of another report.

Unsupervised hierarchical clustering also divided the 25 cancers into 2 groups that coincided almost perfectly with the MMR status (Figure 1). This finding was confirmed by principal component analysis of the colon cancer data. As shown in Figure 2*A* and *B*, component 1, which explained 16% of the variance in this data set, clearly discriminated between MMR-deficient and MMR-proficient cancers. Component 2, which accounted for 11% of all variance, showed no correlation with MMR status. A preliminary investigation suggests that it may be related to the lymph

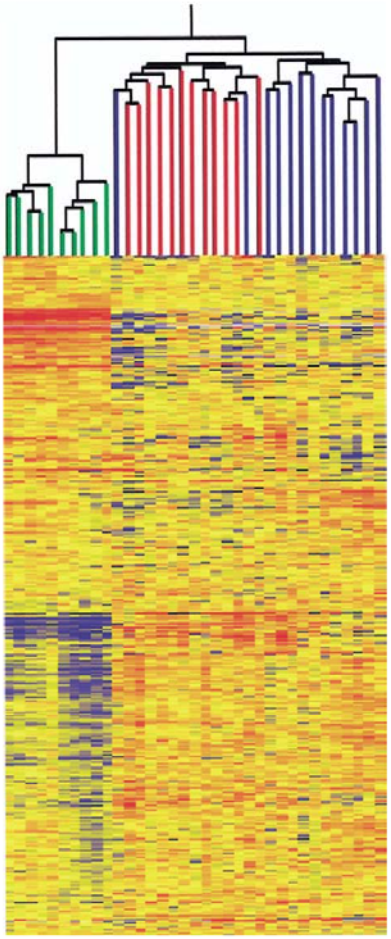


Figure 1. Unsupervised hierarchical clustering of normal colon mucosa and colon cancer samples based on analysis with the Affymetrix gene set after exclusion of unexpressed genes (see supplemental material online at www.gastrojournal.org). The 5858 probes plotted on the y-axis are color coded to indicate their expression levels relative to the median level for the gene across the entire sample set (*blue*, lower; *red*, higher). The 35 tissue samples represented on the x-axis include 10 normal mucosa samples (*green bars*), 12 MMR-deficient cancers (*red bars*), and 13 MMR-proficient cancers (*blue bars*).

node status of the cancers (Figure 2*C*) and the growth patterns of the cancers ("infiltrating" vs "expansive"[13]) (data not shown).

Next, in a supervised analysis based on the MMR status of the cancers, we identified 100 genes that were differentially expressed in MMR-deficient and MMR-proficient cancers with fold differences of ≥2.0 (Table 3). In supervised clustering analysis (Figure 3*A*), this gene set effectively differentiated the 2 tumor subsets and was therefore designated the MMR status signature. It reflects the changes in gene expression (several already identified by others[14,15]) most likely to explain the clinicopathologic differences between MMR-proficient and MMR-deficient colon cancers.

Our next goal was to identify the nature of these expression changes with respect to the normal mucosa and, using the procedure described in Figure 4, to differentiate alterations that are specific to MMR-deficient versus MMR-proficient colon cancers. This process identified 39 genes whose alterations are likely to contribute to tumorigenesis within a specific MMR context. Analysis of these MMR-related tumorigenesis signature genes, which appear in bold in Table 3, reveals down-regulation of 2 putative tumor suppressors (guanylate cyclase 2C,[16] PURA[17]) and up-regulation of 4 putative oncogenes (thymidylate synthase,[18] DUSP4,[19] uPAR,[20] COPS3[21]) as possible markers of MMR-deficient colon cancer tumorigenesis.

In an attempt to isolate changes occurring in the epithelial cell component of the cancers, we compared the colon cancer expression profiles with those of the 14 colon cancer epithelial cell lines. When the relation between expression levels of a gene in MMR-deficient and MMR-proficient cancers (ie, MMR-proficient expression greater than MMR-deficient expression or vice versa) was also observed in the comparison of MMR-deficient and MMR-proficient cell lines, the difference was presumed to reflect an epithelial cell–specific alteration.

Principal component analysis revealed that MMR status also had a substantial impact on the transcriptomes of the colon cancer cell lines (Figure 2*D* and *E*). Supervised statistical group comparison analysis then revealed 66 genes whose transcript levels in MMR-proficient and MMR-deficient cell lines were significantly different, with fold differences of ≥2.0 (Figure 3*B* and Supplementary Table 2, see supplemental material online at www.gastrojournal.org). This set represents a cell line–related version of the MMR status signature identified for colon cancers (Table 3). However, the overlap between these 2 signatures was quite limited (the 12 genes underlined in Table 3) because both had been identified with a fold-difference cutoff of 2.0. Elimination of this cutoff expanded the overlap list to 54 genes, whose altered expression in MMR-defined subgroups of colon cancer cell lines was statistically significant and directionally identical to those observed in the cancers themselves. This set was designated the MMR-related epithelial cell signature (Table 4 and Figure 3*C*).

Some of the differences noted thus far might conceivably reflect transcriptional effects of MSI within the coding or regulatory regions of genes expressed by cancerous epithelial cells. This hypothesis was confirmed by

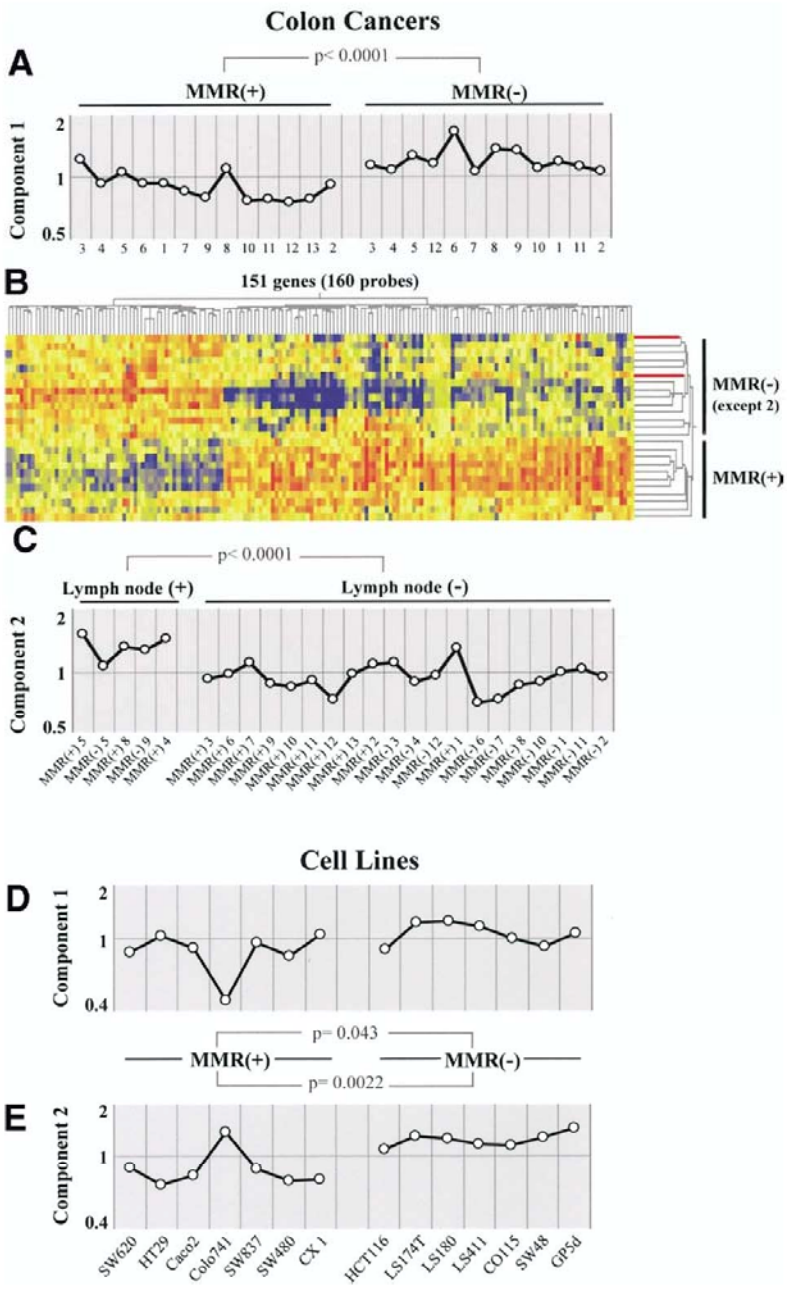


Figure 2. Principal component analysis of the 25 proximal colon cancers and the 14 colon cancer cell lines. (See supplemental material online at www.gastrojournal.org for details on principal component analysis methodology.) (*A*) Normalized component-1 output values (y-axis, log-scale) for the 25 cancer specimens (x-axis, numbered as in Table 1). The difference between values for MMR-proficient and MMR-deficient cancers was highly significant ($t = 4.70$; $P < .0001$). (*B*) Unsupervised hierarchical clustering of the 25 cancers based on the component-1 gene set, that is, the 151 genes (160 probes) with significant principal component analysis values (greater than 0.7 and less than −0.7) in component 1. With 2 exceptions, MMR-proficient cancers 3 and 8 (*red bars*), the cancers formed 2 clusters that coincided with MMR-status subgroups. (*C*) Normalized component-2 output values (y-axis, log-scale) for the 25 cancers (x-axis, numbered as in Table 1). There was a highly significant difference between values for the lymph node–positive and lymph node–negative cancers ($t = 5.27$; $P < .0001$). (*D*) Component-1 and (*E*) component-2 output values (y-axes, log-scale) for the 14 colon cancer cell lines (x-axes). Components 1 and 2 accounted for 20% and 13%, respectively, of the total variance in the cell line data set. In both cases, MMR-deficient and MMR-proficient cell lines were distinguished by significantly different output values (component 1: $t = 2.26$, $P = .043$; component 2: $t = 3.88$, $P = .0022$).

the finding of a higher frequency of microsatellites in the complementary DNAs of the MMR-related epithelial cell signature genes (46%) compared with that (30%) of 100 randomly selected genes expressed without MMR-related variations by the cancers and cell lines (odds ratio, 2.01; 95% confidence interval, 1.01–3.99; $P = .05$). The signature genes also presented significantly longer mononucleotide repeats in the 3′ untranslated regions (UTRs) of microsatellite-containing genes ($t = 3.26$; $P = .0026$) (Table 5). All but one of the MMR-related epithelial cell signature genes containing coding region and/or 5′-UTR mononucleotide repeats displayed decreased transcript

Table 3. *MMR*-Status Signature Genes: 100 Genes Found to Be Differentially Expressed in MMR(+) and MMR(−) Proximal Colon Cancers

Category	Affymetrix ID	GeneBank	Gene name	Gene description	Fold differences[a] C(+) *vs* C(−)	C(−) *vs* NM	C(+) *vs* NM
GENES MORE HIGHLY EXPRESSED IN MMR(+) COLON CANCERS							
Metabolism	1371_s_at	M29874	***CYP2B6***	**Cytochrome P450-IIB**	6.1	−3.7	
	41706_at	AJ130733	*AMACR*	A-methylacyl-CoA racemase	4.7		
	32972_at	Z83819	*NOX1*	Nicotinamide adenine dinucleotide phosphate oxidase 1	4.6		
	37978_at	D78177	*QPRT*	Quinolinate phosphoribosyltransferase	4.3		
	35345_at	X83618	*HMGCS2*	Coenzyme A synthase	3.8		
	35831_at	AB014511	*ATP9A*	Adenosine triphosphatase class II type 9A	3.4		
	36779_at	X90908	*FABP6*	Fatty acid binding protein 6	2.9		
	41352_at	X62822	*SIAT1*	Sialyltransferase 1	2.8		
	40376_at	X83573	*ARSE*	Arylsulfatase E	2.8		
	35966_at	X71125	*QPCT*	Glutamine cyclotransferase	2.6		
	37263_at	U55206	***GGH***	**γ-glutamyl hydrolase**	2.6		2.3
	33936_at	D86181	*GALC*	Galactosylceramidase	2.5		
	35194_at	X53463	***GPX2***	**Glutathione peroxidase 2**	2.5		2.2
	34719_at	AB020645	*GLS*	Glutaminase	2.2		
Extracellular signaling	38582_at	AI961220	*SPINK1*	Serine protease inhibitor	4.6		
	39666_at	U31382	*GNG4*	Guanine nucleotide binding protein	4		
	32649_at	X59871	***TCF7/TCF1***	**Transcription factor 7**	3.5		4.2
	36618_g_at	X77956	***ID1***	**Inhibitor of DNA binding 1**	3.6	−4.2	
	41356_at	W27619	*BCL11A*	B-cell, CLL/lymphoma 11A	3.2		
	34280_at	Y09765	*GABRE*	GABAA receptor epsilon	3.1		
	32207_at	M64925	*MPP1*	Membrane protein, palmitoylated 1	2.9		
	41073_at	AI743745	***GPR49***	**G protein–coupled receptor 49**	2.9		14.6
	34618_at	L38517	*IHH*	Indian hedgehog homolog	2.7		
	35980_at	AB011153	***PLCB1***	**Phospholipase C β 1**	2.7		4
	1291_s_at	L03840	*FGFR4*	Fibroblast growth factor receptor 4	2.7		
	35769_at	AJ011001	***GPR56***	**Epidermal growth factor-TM7–like protein**	2.4		3.3
	39824_at	AI391564	***PTP4A3***	**Protein tyrosine phosphatase**	2.4		3.1
	34450_at	M73489	***GUCY2C***	**Guanylate cyclase 2C**	2.4	−3.1	
	39361_f_at	AF043906	*TM4SF6*	Transmembr 4 superfamily member 6	2.4		
	1603_g_at	L33881	*PRKCI*	Protein kinase C, iota	2.1		
	33213_g_at	AF006751	*RRBP1*	Ribosome binding protein 1	2.1		
	32182_at	AB023182	*NDR2*	Serine/threonine kinase 38 like	2.1		
	35348_at	AF022116	*PRKAB1*	AMP-activated protein kinase	2		
Transcription	31355_at	U77629	***ASCL2***	**Achaete-scute complex-like 2**	3.3		17.4
	36957_at	W22296	*RACK-7*	Protein kinase C binding protein 1	3.1		
	32588_s_at	X78992	***ERF-2***	**Zinc finger protein 36**	2.6	−2.4	
	32583_at	J04111	*JUN*	c-jun proto-oncogene	2.5		
	40061_at	D83784	*PLAGL2*	Pleiomorphic adenoma gene-like 2	2.4		
	41713_at	U09848	*ZNF36*	Zinc finger protein 36	2.3		
	36899_at	M97287	*SATB1*	AT-rich sequence binding protein 1	2.1		
	34195_at	AI121073	*ODAG*	Ocular development–associated	2.1		
	34217_at	AB015132	*KLF7*	Kruppel-like factor	2		
	35614_at	AB012124	***TCFL5***	**Transcription factor–like 5**	2		5.3
Nuclear	34713_at	AB002357	*KIF3B*	Kinesin family member 3B	2.2		
	818_s_at	U72936	***ATRX***	**RAD54 homolog, S cerevisiae**	2.1		2.1
	35221_at	X91648	***PURA***	**Purine-rich binding protein A**	2	−2.1	
	39035_at	AF006010	*DD5*	Progestin induced protein	2.1		
	33845_at	W28483	***HNRPH1***	**Heterogeneous ribonucleoprotein H1**	2.5	−2.1	
Structural	35415_at	X12901	*VIL1*	Villin 1	2.5		
	33917_at	AB002336	*EPB41L1*	Erythrocyte membr prot band 4.1-like 1	2.2		
	32148_at	AI701049	*FARP1*	P63RhoGEF	2.2		
	35362_at	AB018342	*MYO10*	Myosin X	2.2		
	40020_at	AB011536	*CELSR3*	Cadherin seven-pass G-type rec. 3	2.1		
	36061_at	AF009314	***FLJ12815***	**Similar to semaphorin 5A**	2.1	−2.6	
	39598_at	X04325	*GJB1*	Connexin 32	2		
Transmembrane	41188_at	W28186	***LAPTM4B***	**Lysosomal associated protein**	3.1		3.8
	35906_at	L29339	***SLC5A1***	**Na^+ glucose cotransporter**	2.7		2.5

Table 3. (continued).

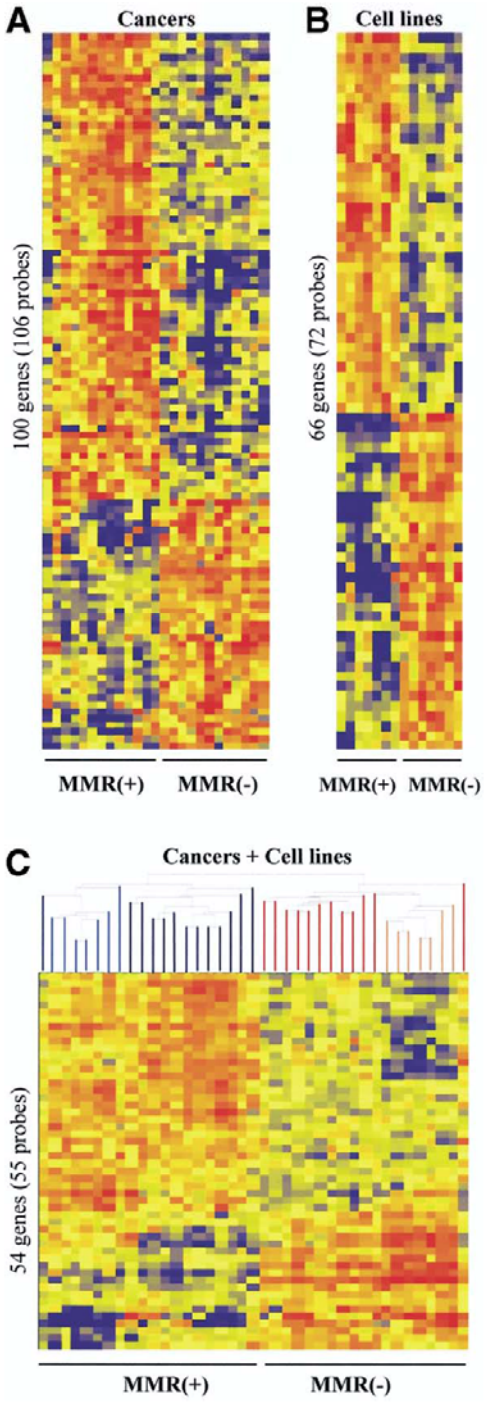
Category	Affymetrix ID	GeneBank	Gene name	Gene description	Fold differences[a] C(+) *vs* C(−)	C(−) *vs* NM	C(+) *vs* NM
	35164_at	AF084481	***WFS1***	**Wolfram syndrome 1**	2.5		4
	37413_at	J05257	***DPEP1***	**Dipeptidase 1**	2.4		5.8
	40885_s_at	N30151	*STX16*	Syntaxin 16	2.3		
Trafficking	34484_at	AI961669	*ARFGEF2*	Guanine nucleotide-exchange factor 2	2.1		
	37751_at	D87444	*TM9SF4*	Transmember 9 superfamily member 4	2		
Unknown	40951_at	AL049250	*KIAA0220*	KIAA0220 protein	3.4		
	38043_at	X55448	*FAM3A*		2.2		
	32502_at	AL041124	***PP1665***	**Hypothetical protein**	2.1		3.8
GENES MORE HIGHLY EXPRESSED IN MMR(−) COLON CANCERS							
Nuclear	35201_at	X16135	***HNRPL***	**Heterog. nuclear ribonucleoprotein L**	3.6	3.6	
	31858_at	X07315	***NUTF2***	**Nuclear transport factor 2**	3.2	8.7	
	39959_at	AL031983	***UBD***	**Ubiquitin D**	2.6	16.3	
	40348_s_at	W25866	***ANP32E***	**Acidic nuclear phosphoprotein 32 E**	2.2	4.2	
	37899_at	X02308	***TYMS***	**Thymidylate synthase**	2.1	2.8	
Signaling	1788_s_at	U48807	***DUSP4***	**Mitogen-activated protein kinase phosphatase 2**	4.1	26.2	
	36014_at	AL033377	*GPR126*	G protein–coupled receptor 126	2.8		
	38489_at	M60047	*FGFBP1*	Fibroblast growth factor binding protein 1	2.3		
	656_at	L08488	*INPP1*	Inositol polyphosphate 1-phosphatase	2.1		
	41169_at	X74039	***uPAR***	**Urokinase plasminogen activator rec**	2.1	2.9	
	33232_at	AI017574	*CRIP1*	Cysteine-rich protein 1	2.1		
	35349_at	AF031647	***COPS3***	**COP9 complex subunit 3**	2.1	2.1	
	960_g_at	HG2463	*CDC42*	Gvanosine triphosphate binding protein, 25 kilodaltons	2		
Apoptosis	38871_at	AJ006288	*BCL10*	B-cell CLL/lymphoma 10	2.4		
	998_s_at	X59770	***IL1R2***	**Interleukin-1 receptor, type II**	2.3		−3.8
	31874_at	Y07846	*GAS2L1*	Growth arrest–specific 2 like 1	2.1		
	1240_at	U13022	*CASP2*	Apoptosis-related protease	2.1		
Immune response	35000_at	U03398	***4-1 BBL***	**Tumor necrosis factor superfamily member 9**	2.4	33.8	
	35937_at	U65416	***MICB***	**Major histocompatibility complex class I chain-related gene B**	2.3	3.6	
	40757_at	M18737	*GZMA*	Granzyme A	2.3		
	37200_at	J04162	*FCGR3A*	Immunoglobulin Fc receptor IIIa gene	2.2		
	37220_at	M63835	***FCGR1A***	**Immunoglobulin Fc receptor Ia gene**	2.1	6.5	
Extracellular	1482_g_at	L23808	*MMP12*	Matrix metalloproteinase 12	2.7		
	2092_s_at	J04765	*SPP1*	Osteopontin	2.4		
Transcription	37141_at	U39840	*FOXA1*	Hepatocyte nuclear factor-3α	2.7		
	869_at	U14193	*GTF2A2*	General transcription factor IIA	2.1		
Others	36130_f_at	R92331	***MT1E***	**Metallothionein 1E**	2.5		−5.6
	31623_f_at	K01383	***MT1A***	**Metallothionein I A**	2.6		−5
	34775_at	AF065388	*TSPAN-1*	Tetraspan 1	2.4		
	847_at	U17969	***EIF5A***	**Euk translation initiation factor 5A**	2.4	2.8	
	37976_at	AL034397	*VSIG4*	V-set and Ig domain containing 4	2.4		
	37552_at	U33632	*KCNK1*	Potassium channel	2.3		
	37351_at	X90858	*UPP1*	Uridine phosphorylase 1	2.1		
	717_at	D87119	***TRIB2***	**Tribbles homolog 2**	2	5	
Unknown	38250_at	D26488	***KIAA0007***	**KIAA0007 protein**	2.7	4.7	

NOTE. Bold type indicates *MMR-related tumorigenesis signature* genes (see Figure 4). Thirty-nine genes whose altered expression (vs normal mucosa) were observed exclusively in MMR(+) OR MMR(−) colon cancers. The fold changes with respect to normal mucosa are shown in the last 2 columns.

Underlined type indicates 12 genes displaying similar patterns of differential expression [MMR(+) vs. MMR(−)] in both colon cancers and colon cancer cell lines (no fold change restriction was applied to cell lines).

C(+), MMR-proficient cancer; C(−), MMR-deficient cancer; NM, normal mucosa; MMR(+), MMR proficient; MMR(−), MMR deficient.

[a]Statistical significance was assessed with the Mann–Whitney test with a false discovery rate of 0.05 and Benjamini–Hochberg multiple testing correction. Minimal fold difference, 2.0: genes whose differential expression was characterized by fold differences <2.0 are not shown, even if statistically significant.



levels in MMR-deficient cells, whereas those with 3′-UTR microsatellites presented higher messenger RNA levels in these cells. The cell line expression of 3 genes from this latter group (*NUTF2*, *HNRPL*, and *CTNNB1*) was analyzed with quantitative reverse-transcription polymerase chain reaction. Their coding regions were amplified with similar efficiency in all 14 lines, but amplification of the repeat-containing 3′-UTRs was much more efficient in the MMR-deficient cell lines (Supplementary Figure 1, see supplemental material online at www.gastrojournal.org). Moreover, in the MMR-deficient lines, the long 3′-UTR mononucleotide stretches present in these 3 genes were always several nucleotides shorter than their counterparts in the MMR-proficient lines (data not shown), suggesting that the more efficient 3′-UTR amplification was somehow related to MSI.

We then focused our analysis on *4-1BBL*, a gene involved in the antitumoral immune response. Levels of RNA for this gene in MMR-deficient cancers and cell lines were, respectively, 2.4 and 6.0 times higher than those observed in their MMR-proficient counterparts (Figure 5*A*). This difference was confirmed by quantitative reverse-transcription polymerase chain reaction amplification of the 4-1BBL coding region (Figure 5*B*) and flow cytometric assessment of 4-1BBL levels (Figure 5*C*) in colon cancer cell lines.

Discussion

Analysis of the MMR status and MMR-related epithelial cell signatures shows several changes that might explain the better prognosis associated with MMR-deficient colon cancers. Compared with MMR-proficient cancers, MMR-deficient tumors presented reduced transcript levels of several genes with functions that presumably favor tumor progression (eg, *PTP4A3*,[22] *PRKCI*,[23,24] *VEGFB*,[25] *B-myb*,[26] *SAS*,[27] *SIAT1*,[28] *NOX1*,[29] *DD5*,[30] and *IHH*[31]) and increased expression of factors believed to inhibit tumor growth, progression, and/or invasiveness. The latter included the proapoptotic

Figure 3. Supervised hierarchical clustering of colon cancers and colon cancer cell lines classified on the basis of their MMR status. (*A*) Clustering of 25 colon cancers by the MMR-status signature gene set (100 genes, 106 probes) shown in Table 3. (*B*) Clustering of 14 colon cancer cell lines by the 66 genes shown in Supplementary Table 2, which can be considered a cell line–specific version of the MMR-status signature set. (*C*) Clustering of colon cancers and cell lines by the MMR-related epithelial cell signature gene set shown in Table 4. Sample tree: MMR-proficient cell lines in light blue, MMR-proficient cancers in dark blue, MMR-deficient cell lines in orange, and MMR-deficient cancers in red.

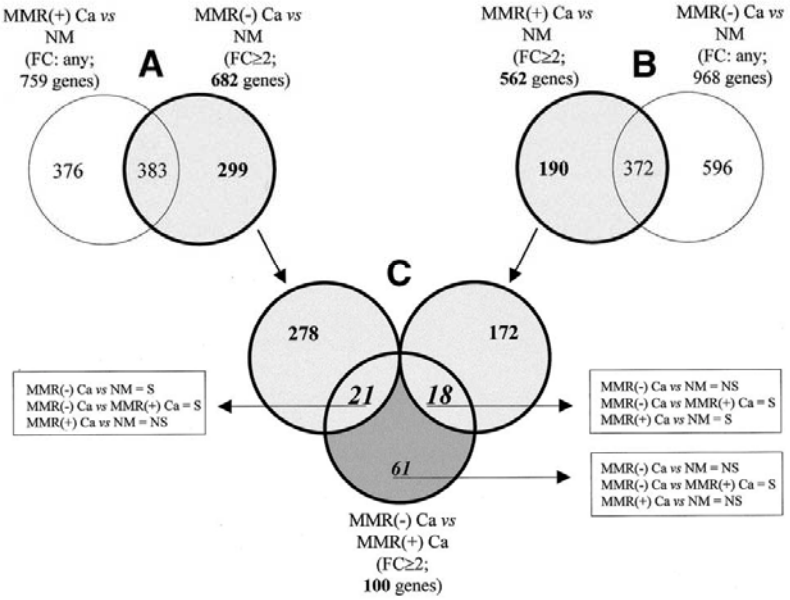


Figure 4. Selection of the MMR-related tumorigenesis signature gene set from the MMR-status signature gene set. (*A*) The *light gray circle* represents the 682 genes whose expression levels in MMR-deficient cancers were significantly increased or decreased, relative to those in normal mucosa (NM), with fold changes (FC) of ≥2.0. (The 10 normal mucosa samples were considered as a homogeneous group because their variability was much less marked than that observed among the cancer specimens, as shown in Figure 1). The *white circle* represents 759 genes displaying significantly altered expression (no FC restrictions) in MMR-proficient cancers with respect to normal mucosa. (Elimination of the FC cutoff in this selection allowed us to identify MMR-deficient–specific changes with greater accuracy.) The intersection of the 2 circles represents the 383 genes displaying similar alterations in MMR-deficient and MMR-proficient cancers with respect to normal mucosa. Exclusion of these genes leaves 299 genes whose expression was altered (relative to normal mucosa) exclusively in MMR-deficient cancers. (*B*) A similar procedure was used to identify 190 genes with significant expression changes (relative to normal mucosa) that were confined exclusively to MMR-proficient cancers. (*C*) The *light gray circles* here represent the sets of 299 and 190 genes obtained in *A* and *B*, while the *dark gray circle* represents the 100 genes making up the MMR-status signature (Table 3). Gene set characteristics in terms of differences between transcriptional levels in MMR-deficient cancers, MMR-proficient cancers, and normal mucosa are summarized in the small rectangles; differences are classified as significant (S) or nonsignificant (NS). Overlap between the light gray circle on the left and the dark gray circle represents 21 genes whose expression levels in MMR-deficient cancers were significantly different (FC ≥2.0) from those observed in normal mucosa and also significantly different from those of MMR-proficient cancers. Intersection of the dark gray circle and the light gray circle on the right defines 18 other genes with expression levels in MMR-proficient cancers that differed significantly (FC ≥2.0) from those of normal mucosa and also from those of MMR-deficient cancers. These 39 genes represent a subset of the MMR-status signature gene set, which we refer to as the MMR-related tumorigenesis signature.

genes *caspase 2*,[32] *BCL10*,[33] *IL1RII*,[34] and *GAS2*L1[35]; the transforming growth factor β effector *SMAD4*[36]; the matrix metalloproteinase *HME*[37]; and *SPP1*, which mediates the tumor suppressor function of p53 by activating host immunosurveillance.[38] The reportedly lower frequency of p53 mutations in MMR-deficient colon cancers[39] is consistent with evidence of enhanced transcription of several p53-induced genes (eg, $p21^{Waf1/Cip1}$,[40] *Maspin*,[41] and *DR-NM23*[42]) in our MMR-deficient cancers and cell lines (Supplementary Table 2, see supplemental material online at www.gastrojournal.org). The nonsignificance of this difference in the cancer specimens is probably due to the presence of wild-type *p53* in stromal cells, which would attenuate differences related to the higher frequency of epithelial cell *p53* mutations in MMR-proficient cancers.

Other genes displaying higher expression in MMR-deficient cells are involved in the antitumoral immune response, including *4-1BBL*,[43] *WSX-1*,[44] *MICB*,[45] *FCGR3A*,[46] and *granzyme A*.[47] In contrast, the granzyme A inhibitor, *SPINK1*,[48] was markedly underexpressed in MMR-deficient cancers, an alteration that should enhance the antitumoral effects of natural killer cells and cytotoxic T lymphocytes.

Overexpression of *4-1BBL* was also documented at the protein levels. This gene is constitutively expressed by activated antigen-presenting cells (eg, dendritic cells, B cells, macrophages) but also by several leukemia and carcinoma cells.[49] Interaction of *4-1BBL* with the tumor necrosis factor receptor 4-1BB plays a crucial role in the response of memory CD8 T cells to tumor cells[43,50] and is an important determinant of effective antitumoral

Table 4. MMR-Related Epithelial Cell Signature Genes: Genes Displaying Similar Differential Expression Patterns [MMR(+) Versus MMR(−) in Both Colon Cancers and Colon Cancer Cell Lines

					Fold differences: MMR(+) *vs* MMR(−)	
Category	Affymetrix ID	GeneBank	Gene name	Gene description	Cancers	Cell lines
GENES MORE HIGHLY EXPRESSED IN MMR(+) CANCERS AND CELL LINES[a]						
Signaling	1603_g_at	L33881	*PRKCI*	Protein kinase C, iota	2.1	2.1
	1854_at	X13293	*B-myb*	Myeloblastosis viral oncogene	1.9	1.9
	36083_at	U01160	*SAS*	Sarcoma amplified sequence	1.8	2.1
	1926_at	U48801	*VEGFB*	Vascular endothelial growth factor B	1.4	2.2
	33456_at	U24166	*EB1*	Microtubule-associated protein	1.9	1.9
	32324_at	X57346	*14-3-3 beta*	tyr 3/trp 5-monooxygenase activator	1.7	2
	41220_at	AB023208	*MSF*	MLL septin-like fusion	1.3	1.6
	31536_at	AB020693	*RTN4/nogo-B*	Reticulon 4	1.4	2.1
	41134_at	AB023181	*DLGAP4*	Discs large homolog-associated 4	1.3	1.4
	1814_at	D50683	*TGFbetaRII*	Transforming growth factor β receptor II	1.8	2.6
Transcription	40061_at	D83784	*PLAGL2*	Pleiomorphic adenoma gene-like 2	2.4	1.7
	34195_at	AL121073	*ODAG*	Ocular development-associated	2.1	2.4
	34396_at	AB023195	*ASXL1*	Additional sex combs like 1	1.9	2.6
	34868_at	AB029012	*SMG-5*	Est 1p-like protein B	1.3	1.9
	40411_at	D80003	*NCOA6/ASC2*	Nuclear receptor coactivator 6	1.7	1.6
Nuclear	35221_at	X91648	*PURA*	Purine-rich binding protein A	2	2.4
	33845_at	W28483	*HNRPH1*	Heterogeneous ribonucleoprotein H1	2.5	4
	36125_s_at	L38696	*RALY*	RNA binding protein	1.8	2.4
	35682_at	AI133727	*ZC3HAV1*	Zinc finger CCCH type, antiviral 1	1.4	2
Structural	40725_at	AF047438	*GOSR1*	Golgi SNAP receptor complex 1	1.6	2.2
	41757_at	W25933	*VAPB*	Vesicle-associated membrane protein	1.7	2.2
	34031_i_at	U90269	*CCM1*	Cerebral cavernous malformations 1	1.4	1.6
Transporters	37751_at	D87444	*TM9SF4*	Transmembr 9 superfamily member 4	2	1.8
	39523_at	AF038897	*STX16*	Syntaxin 16	1.9	1.7
RNA binding	41823_at	AJ132258	*STAU*	Staufen, RNA binding protein	1.9	1.7
	34345_at	AF026031	*C20orf14*	Chrom 20 open-reading frame 14	1.8	1.8
	32758_g_at	U84720	*RAE1*	RAE1 RNA export 1 homolog	1.7	1.7
Others	39005_s_at	AB018257	*ZNF294*	Zinc finger protein 294	1.7	2.1
	37410_at	AJ224358	*SURF5*	Surfeit 5	1.7	1.7
	33418_at	AL096752	*RAB3GAP*	RAB3 guanosine triphosphatase–activating protein	1.5	1.7
	39429_at	X99050	*UVRAG*	UV radiation resistance associated	1.3	1.6
Unknown function	40517_at	AB002370	*KIAA0372*	KIAA0372 protein	1.7	3.8
	38204_at	AB007866	*KIAA0406*	KIAA0406 protein	1.7	1.5
	33242_at	U92980	*DT1P1A10*	Hypothetical protein	1.6	1.7
GENES MORE HIGHLY EXPRESSED IN MMR(−) CANCERS AND CELL LINES[a]						
Nuclear	35201_at	X16135	*HNRPL*	Heterogeneous nuclear ribonucleoprot L	3.6	6.5
	31858_at	X07315	*NUTF2*	Nuclear transport factor 2	3.2	3.7
	40348_s_at	W25866	*ANP32E*	Acidic nuclear phosphoprotein 32 E	2.2	2.1
	35553_at	AF042181	*TSPYL1*	Testis-specific Y-encoded-like protein	1.8	4.4
Signaling	38489_at	M60047	*FGFBP1*	Fibroblast growth factor binding protein 1	2.3	11.1
	40777_at	X87838	*CTNNB1*	β-catenin	1.7	2.6
	36953_at	U44378	*SMAD4*	SMAD, mothers against DPP homolog 4	1.5	2.7
	33770_at	AF009225	*IKKα*	Iκ B kinase α	1.3	1.6
Apoptosis	1240_at	U13022	*CASP2*	Apoptosis-related protease	2.1	2.2
	38871_at	AJ006288	*BCL10*	B-cell CLL/lymphoma 10	2.4	2.9
Immune response	35000_at	U03398	*4-1 BBL*	Tumor necrosis factor superfamily member 9	2.4	6
	37844_at	AI263885	*WSX-1*	Interleukin-27 receptor, alpha	1.9	2.6
Metabolism	37992_s_at	AI436567	*ATP5D*	Adenosine triphosphate synthase delta subunit	1.7	2
	36599_at	M55905	*ME2*	Malic enzyme 2	1.5	2.7
	41241_at	D84273	*NARS*	Asparaginyl-tRNA synthetase	1.4	1.6
	38485_at	AA760866	*NDUFC1*	Nicotinamide adenine dinucleotide dehydrogenase	1.4	1.5
	40096_at	D14710	*ATP5A1*	Adenosine triphosphate synthase α subunit	1.3	1.6
RNA binding	32440_at	X53777	*RPL17*	Ribosomal protein L17	1.3	1.4
Unknown function	33844_at	AA160724		Similar to RCD1 homolog 1	1.7	2.3
	38250_at	D26488	*KIAA0007*	KIAA0007 protein	2.7	3.1

MMR(+), MMR proficient; MMR(−), MMR deficient.

[a]Statistically significant MMR-related differences observed in cancers and those observed in cell lines (both assessed with the Mann–Whitney test with a false discovery rate of 0.05 and Benjamini–Hochberg multiple testing correction).

immune responses.[43,51] Its increased expression in MMR-deficient colon cancer cells might thus represent a molecular link between the intense lymphocytic infiltration and relatively good prognosis that characterize these tumors.

MSI could conceivably be responsible for the decreased transcript levels of genes containing 5′-UTR or coding-

Table 5. *MMR-Related Epithelial Cell Signature* Genes Harboring Mononucleotide Repeats in Their Coding Region (CDS) or Untranslated Regions (5′UTR or 3′UTR)

Gene name[a]	Repeat (cDNA)	mRNA region	MMR(−) versus MMR(+)[a]
KIAA0372	(T)8	5′UTR	Decreased
STAU	(C)9	5′UTR	Decreased[b]
VAPB	(C)9	5′UTR	Decreased[b]
ASXL1	(C)8	5′UTR	Decreased[b]
	(G)8	CDS	
HNRPH1	(T)10	5′UTR	Decreased
	(T)8	CDS	
	(T)13, (T)19	3′UTR	
TGFβRII	(A)10	CDS	Decreased
PRKCI	(A)8	CDS	Decreased
ZNF294	(A)11	CDS	Decreased
UVRAG	(A)10	CDS	Decreased
RAE1	(A)8	CDS	Decreased[b]
FGFBP1	(A)9	CDS	Increased
RTN4/Nogo-B	(A)9	CDS	Decreased
	(A)16	3′UTR	
CCM1	(A)8	CDS	Decreased
	(A)15	3′UTR	
HNRPL	(C)8	CDS	Increased
	(T)21	3′UTR	
CTNNB1	(T)18, (T)28	3′UTR	Increased
BCL-10	(T)14	3′UTR	Increased
CASP2	(A)19, (T)25	3′UTR	Increased
TSPYL1	(T)26	3′UTR	Increased
WSX-1	(A)16	3′UTR	Increased
ANP32E	(T)20	3′UTR	Increased
KIAA0007	(T)23	3′UTR	Increased
NUTF2	(T)22	3′UTR	Increased
SAS	(T)13	3′UTR	Decreased
14-3-3 β	(A)23	3′UTR	Decreased
ZC3HAV1	(T)23, (A)30	3′UTR	Decreased

NOTE. In all 154 genes analyzed, mononucleotide repeats in the 3′UTRs were much longer that those in the 5′UTRs and coding regions, which is consistent with a previous observation (Park et al, Cancer Res 2002;62:1284). We therefore limited our search to coding-region and 5′UTR microsatellites consisting of ≥8 repeat units and 3′UTR microsatellites with ≥12 repeat units.

MMR(+), MMR proficient; MMR(−), MMR deficient.

[a]To our knowledge, only 2 of these genes have previously been identified as MSI targets: *TGFβRII*[6] and *UVRAG*[50] (the latter based on the analysis of cDNA microarray data). Several known MSI targets (*BAX, MSH3, RACK7, CBF2, RECQL1,* and *MRE11*) also displayed MMR(−)-related down-regulation in our study, but they are not included in this table because the decreases were not statistically significant in both the colon cancers and colon cancer cell lines.

[b]Transcript levels in MMR(−) cancers and MMR(−) cell lines *versus* expression levels in their MMR(+) counterparts. MMR(+) expression levels were not significantly different from those observed in normal mucosal samples except for 4 genes as labeled with a footnote; in these cases, the decreased expression in MMR(−) cancers and cell lines reflects an MMR(+)-related increase over normal mucosal levels.

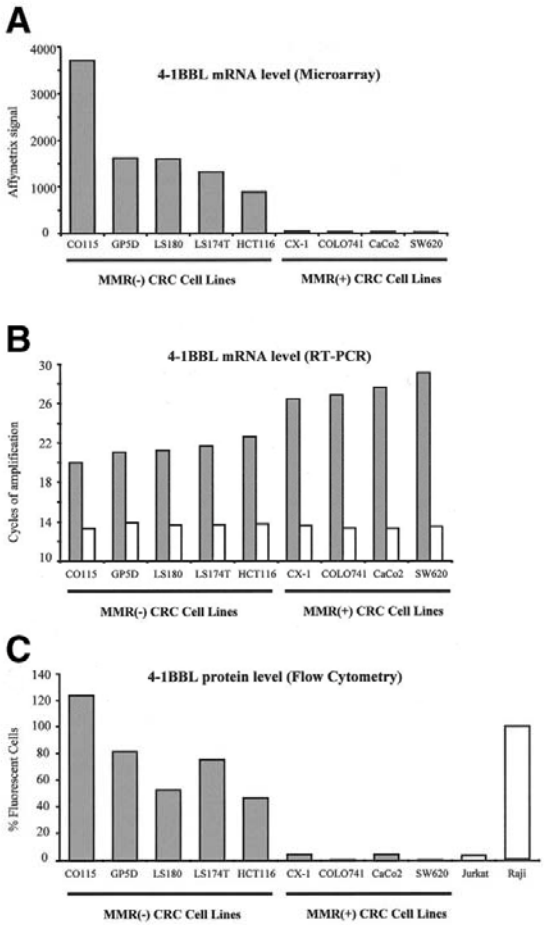


Figure 5. 4-1BBL messenger RNA (mRNA) and protein expression in MMR-deficient and MMR-proficient colon cancer cell lines. (No data are shown for 5 of the cell lines, which could not be reliably analyzed with flow cytometry.) (*A*) *Gray bars* represent mRNA levels based on raw signals (y-axis) detected in MMR-deficient (x-axis; left) and MMR-proficient (x-axis; right) cell lines with the Affymetrix U95v2 microarray. (*B*) *Gray bars* represent mRNA levels for the *4-1BBL* coding region in MMR-deficient (*left*) and MMR-proficient (*right*) cell lines determined by quantitative reverse-transcription polymerase chain reaction (RT-PCR); *white bars* represent corresponding levels for amplification of the control housekeeping gene *GAPDH*. The *4-1BBL* coding region was amplified with primers complementary to the exon 2/exon 3 boundary (forward) and to the central region of exon 3 (reverse). The number of cycles (y-axis) at the beginning of the log-phase amplification is inversely correlated with the abundance of the transcript. (*C*) *Gray bars* represent cytoplasmic membrane expression of 4-1BBL polypeptide assessed by flow cytometry. Jurkat and Raji cells (*white bars*) were used as negative and positive controls, respectively, and expression values in colon cancer cell lines were normalized to that of Raji cells. All RT-PCR and flow cytometry assays were performed in duplicate.

region microsatellites (Table 5). Alterations in the 5′-UTR could diminish transcription efficiency, whereas coding-region MSI would lead to nonsense-mediated decay of the RNA transcript.[52] In contrast, genes whose 3′-UTRs contained long mononucleotide repeats generally displayed increased expression in MMR-deficient settings, and the 3 genes of this type tested for MSI all displayed repeat shortening. The fact that in reverse-transcription polymerase chain reaction the microsatellite-containing tracts of the 3′-UTRs of these genes (but not their coding regions) were amplified much more efficiently in MMR-deficient cells (Supplementary Figure 1, see supplemental material online at www.gastrojournal.org) is difficult to explain, but recent evidence suggests that poly(U)-tract length can affect RNA stability and processing and translation efficiency.[53–56]

Appendix

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1053/j.gastro.2005.06.028.

References

1. Marra G, Boland CR. Hereditary nonpolyposis colorectal cancer: the syndrome, the genes, and historical perspectives. J Natl Cancer Inst 1995;87:1114–1125.
2. Herman JG, Umar A, Polyak K, Graff JR, Ahuja N, Issa JP, Markowitz S, Willson JK, Hamilton SR, Kinzler KW, Kane MF, Kolodner RD, Vogelstein B, Kunkel TA, Baylin SB. Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. Proc Natl Acad Sci U S A 1998;95:6870–6875.
3. Marra G, Schär P. Recognition of DNA alterations by the mismatch repair system. Biochem J 1999;338:1–13.
4. Truninger K, Menigatti M, Luz J, Russell A, Haider R, Gebbers J-O, Bannwart F, Yurtsever H, Neuweiler J, Riehle H-M, Cattaruzza MS, Heinimann K, Schär P, Jiricny J, Marra G. Immunohistochemical analysis reveals high frequency of PMS2 defects in colorectal cancer. Gastroenterology 2005;128:1160–1171.
5. Ionov Y, Peinado MA, Malkhosyan S, Shibata D, Perucho M. Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. Nature 1993;363:558–561.
6. Markowitz S, Wang J, Myeroff L, Parsons R, Sun L, Lutterbaugh J, Fan RS, Zborowska E, Kinzler KW, Vogelstein B, et al. Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability. Science 1995;268:1336–1338.
7. Rampino N, Yamamoto H, Ionov Y, Li Y, Sawai H, Reed JC, Perucho M. Somatic frameshift mutations in the BAX gene in colon cancers of the microsatellite mutator phenotype. Science 1997;275:967–969.
8. Jass JR. Familial colorectal cancer: pathology and molecular characteristics. Lancet Oncol 2000;1:220–226.
9. Samowitz WS, Curtin K, Ma KN, Schaffer D, Coleman LW, Leppert M, Slattery ML. Microsatellite instability in sporadic colon cancer is associated with an improved prognosis at the population level. Cancer Epidemiol Biomarkers Prev 2001;10:917–923.
10. Sankila R, Aaltonen LA, Jarvinen HJ, Mecklin JP. Better survival rates in patients with MLH1-associated hereditary colorectal cancer. Gastroenterology 1996;110:682–687.
11. Cejka P, Stojic L, Mojas N, Russell AM, Heinimann K, Cannavo E, di Pietro M, Marra G, Jiricny J. Methylation-induced G(2)/M arrest requires a full complement of the mismatch repair protein hMLH1. EMBO J 2003;22:2245–2254.
12. di Pietro M, Marra G, Cejka P, Stojic L, Menigatti M, Cattaruzza MS, Jiricny J. Mismatch repair-dependent transcriptome changes in human cells treated with the methylating agent N-methyl-n′-nitro-N-nitrosoguanidine. Cancer Res 2003;63:8158–8166.
13. Fgashira Y, Yoshida T, Hirata I, Hamamoto N, Akutagawa H, Takeshita A, Noda N, Kurisu Y, Shibayama Y. Analysis of pathological risk factors for lymph node metastasis of submucosal invasive colon cancer. Mod Pathol 2004;17:503–511.
14. Mori Y, Selaru FM, Sato F, Yin J, Simms LA, Xu Y, Olaru A, Deacu E, Wang S, Taylor JM, Young J, Leggett B, Jass JR, Abraham JM, Shibata D, Meltzer SJ. The impact of microsatellite instability on the molecular phenotype of colorectal tumors. Cancer Res 2003;63:4577–4582.
15. Banerjea A, Ahmed S, Hands RE, Huang F, Han X, Shaw PM, Feakins R, Bustin SA, Dorudi S. Colorectal cancers with microsatellite instability display mRNA expression signatures characteristic of increased immunogenicity. Mol Cancer 2004;3:21.
16. Carrithers SL. Diarrhea or colorectal cancer: can bacterial toxins serve as a treatment for colon cancer? Proc Natl Acad Sci U S A 2003;100:3018–3020.
17. Johnson EM. The Pur protein family: clues to function from recent studies on cancer and AIDS. Anticancer Res 2003;23:2093–2100.
18. Rahman L, Voeller D, Rahman M, Lipkowitz S, Allegra C, Barrett JC, Kaye FJ, Zajac-Kaye M. Thymidylate synthase as an oncogene: a novel role for an essential DNA synthesis enzyme. Cancer Cell 2004;5:341–351.
19. Wang HY, Cheng Z, Malbon CC. Overexpression of mitogen-activated protein kinase phosphatases MKP1, MKP2 in human breast cancer. Cancer Lett 2003;191:229–237.
20. Berger DH. Plasmin/plasminogen system in colorectal cancer. World J Surg 2002;26:767–771.
21. van Dartel M, Hulsebos TJ. Amplification and overexpression of genes in 17p11.2-p12 in osteosarcoma. Cancer Genet Cytogenet 2004;153:77–80.
22. Saha S, Bardelli A, Buckhaults P, Velculescu VE, Rago C, St Croix B, Romans KE, Choti MA, Lengauer C, Kinzler KW, Vogelstein B. A phosphatase associated with metastasis of colorectal cancer. Science 2001;294:1343–1346.
23. Murray NR, Jamieson L, Yu W, Zhang J, Gokmen-Polar Y, Sier D, Anastasiadis P, Gatalica Z, Thompson EA, Fields AP. Protein kinase Ciota is required for Ras transformation and colon carcinogenesis in vivo. J Cell Biol 2004;164:797–802.
24. Zhang J, Anastasiadis PZ, Liu Y, Thompson EA, Fields AP. Protein kinase C (PKC) betaII induces cell invasion through a Ras/Mek-, PKC iota/Rac 1-dependent signaling pathway. J Biol Chem 2004;279:22118–22123.
25. Hanrahan V, Currie MJ, Gunningham SP, Morrin HR, Scott PA, Robinson BA, Fox SB. The angiogenic switch for vascular endothelial growth factor (VEGF)-A, VEGF-B, VEGF-C, and VEGF-D in the adenoma-carcinoma sequence during colorectal cancer progression. J Pathol 2003;200:183–194.
26. Amatschek S, Koenig U, Auer H, Steinlein P, Pacher M, Gruenfelder A, Dekan G, Vogl S, Kubista E, Heider KH, Stratowa C, Schreiber M, Sommergruber W. Tissue-wide expression profiling using cDNA subtraction and microarrays to identify tumor-specific genes. Cancer Res 2004;64:844–856.
27. Arai H, Ueno T, Tangoku A, Yoshino S, Abe T, Kawauchi S, Oga A, Furuya T, Oka M, Sasaki K. Detection of amplified oncogenes by genome DNA microarrays in human primary esophageal squa-

mous cell carcinoma: comparison with conventional comparative genomic hybridization analysis. Cancer Genet Cytogenet 2003;146:16–21.
28. Kemmner W, Kruck D, Schlag P. Different sialyltransferase activities in human colorectal carcinoma cells from surgical specimens detected by specific glycoprotein and glycolipid acceptors. Clin Exp Metastasis 1994;12:245–254.
29. Arbiser JL, Petros J, Klafter R, Govindajaran B, McLaughlin ER, Brown LF, Cohen C, Moses M, Kilroy S, Arnold RS, Lambeth JD. Reactive oxygen generated by Nox1 triggers the angiogenic switch. Proc Natl Acad Sci U S A 2002;99:715–720.
30. Clancy JL, Henderson MJ, Russell AJ, Anderson DW, Bova RJ, Campbell IG, Choong DY, Macdonald GA, Mann GJ, Nolan T, Brady G, Olopade OI, Woollatt E, Davies MJ, Segara D, Hacker NF, Henshall SM, Sutherland RL, Watts CK. EDD, the human orthologue of the hyperplastic discs tumour suppressor gene, is amplified and overexpressed in cancer. Oncogene 2003;22:5070–5081.
31. Qualtrough D, Buda A, Gaffield W, Williams AC, Paraskeva C. Hedgehog signalling in colorectal tumour cells: induction of apoptosis with cyclopamine treatment. Int J Cancer 2004;110:831–837.
32. Norbury CJ, Zhivotovsky B. DNA damage-induced apoptosis. Oncogene 2004;23:2797–2808.
33. Mullauer L, Gruber P, Sebinger D, Buch J, Wohlfart S, Chott A. Mutations in apoptosis genes: a pathogenetic factor for human disease. Mutat Res 2001;488:211–231.
34. Waterhouse CC, Joseph RR, Stadnyk AW. Endogenous IL-1 and type II IL-1 receptor expression modulate anoikis in intestinal epithelial cells. Exp Cell Res 2001;269:109–116.
35. Benetti R, Del Sal G, Monte M, Paroni G, Brancolini C, Schneider C. The death substrate Gas2 binds m-calpain and increases susceptibility to p53-dependent apoptosis. EMBO J 2001;20:2702–2714.
36. Ramachandra M, Atencio I, Rahman A, Vaillancourt M, Zou A, Avanzini J, Wills K, Bookstein R, Shabram P. Restoration of transforming growth factor Beta signaling by functional expression of smad4 induces anoikis. Cancer Res 2002;62:6045–6051.
37. Yang W, Arii S, Gorrin-Rivas MJ, Mori A, Onodera H, Imamura M. Human macrophage metalloelastase gene expression in colorectal carcinoma and its clinicopathologic significance. Cancer 2001;91:1277–1283.
38. Morimoto I, Sasaki Y, Ishida S, Imai K, Tokino T. Identification of the osteopontin gene as a direct target of TP53. Genes Chromosomes Cancer 2002;33:270–278.
39. Jass JR, Whitehall VL, Young J, Leggett BA. Emerging concepts in colorectal neoplasia. Gastroenterology 2002;123:862–876.
40. Coqueret O. New roles for p21 and p27 cell-cycle inhibitors: a function for each cell compartment? Trends Cell Biol 2003;13:65–70.
41. Oshiro MM, Watts GS, Wozniak RJ, Junk DJ, Munoz-Rodriguez JL, Domann FE, Futscher BW. Mutant p53 and aberrant cytosine methylation cooperate to silence gene expression. Oncogene 2003;22:3624–3634.
42. Venturelli D, Martinez R, Melotti P, Casella I, Peschle C, Cucco C, Spampinato G, Darzynkiewicz Z, Calabretta B. Overexpression of DR-nm23, a protein encoded by a member of the nm23 gene family, inhibits granulocyte differentiation and induces apoptosis in 32Dc13 myeloid cells. Proc Natl Acad Sci U S A 1995;92:7435–7439.
43. Cheuk AT, Mufti GJ, Guinn BA. Role of 4-1BB:4-1BB ligand in cancer immunotherapy. Cancer Gene Ther 2004;11:215–226.
44. Hisada M, Kamiya S, Fujita K, Belladonna ML, Aoki T, Koyanagi Y, Mizuguchi J, Yoshimoto T. Potent antitumor activity of interleukin-27. Cancer Res 2004;64:1152–1156.
45. Collins RW. Human MHC class I chain related (MIC) genes: their biological function and relevance to disease and transplantation. Eur J Immunogenet 2004;31:105–114.
46. Clynes RA, Towers TL, Presta LG, Ravetch JV. Inhibitory Fc receptors modulate in vivo cytotoxicity against tumor targets. Nat Med 2000;6:443–446.
47. Lieberman J, Fan Z. Nuclear war: the granzyme A-bomb. Curr Opin Immunol 2003;15:553–559.
48. Tsuzuki S, Kokado Y, Satomi S, Yamasaki Y, Hirayasu H, Iwanaga T, Fushiki T. Purification and identification of a binding protein for pancreatic secretory trypsin inhibitor: a novel role of the inhibitor as an anti-granzyme A. Biochem J 2003;372:227–233.
49. Salih HR, Kosowski SG, Haluska VF, Starling GC, Loo DT, Lee F, Aruffo AA, Trail PA, Kiener PA. Constitutive expression of functional 4-1BB (CD137) ligand on carcinoma cells. J Immunol 2000;165:2903–2910.
50. Wiethe C, Dittmar K, Doan T, Lindenmaier W, Tindle R. Provision of 4-1BB ligand enhances effector and memory CTL responses generated by immunization with dendritic cells expressing a human tumor-associated antigen. J Immunol 2003;170:2912–2922.
51. Martinet O, Ermekova V, Qiao JQ, Sauter B, Mandeli J, Chen L, Chen SH. Immunomodulatory gene therapy with interleukin 12 and 4-1BB ligand: long-term remission of liver metastases in a mouse model. J Natl Cancer Inst 2000;92:931–936.
52. Ionov Y, Nowak N, Perucho M, Markowitz S, Cowell JK. Manipulation of nonsense mediated decay identifies gene mutations in colon cancer cells with microsatellite instability. Oncogene 2004;23:639–645.
53. Wilusz CJ, Wormington M, Peltz SW. The cap-to-tail guide to mRNA turnover. Nat Rev Mol Cell Biol 2001;2:237–246.
54. Beaudoing E, Gautheret D. Identification of alternate polyadenylation sites and analysis of their tissue distribution using EST data. Genome Res 2001;11:1520–1526.
55. Mazumder B, Seshadri V, Fox PL. Translational control by the 3′-UTR: the ends specify the means. Trends Biochem Sci 2003;28:91–98.
56. Ambros V. The functions of animal microRNAs. Nature 2004;431:350–355.

Received March 15, 2005. Accepted June 2, 2005.

Address requests for reprints to: Giancarlo Marra, MD, PhD, Institute of Molecular Cancer Research, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland. e-mail: marra@imcr.unizh.ch; fax: (41) 01 635 34 84.

M.D.P. and J.S.B. contributed equally to this study.

Supported by Swiss National Science Foundation grant 3100-068182 (to M.D.P. and J.J.), the Swiss Cancer League (J.S.B. and A.R.), Hanne-Liebermann Stiftung, Forschung im Gesundheitswesen (K.T.), and EU grant QLG1-CT-2000-01230 (to J.J. and G.M.).

The authors thank Witold Filipowicz and Emilija Veljkovic for critical reading of the manuscript; Ritva Haider, Judith Luz, Beatrix Boesch, Sibel Yeginsoy, and Eva Niederer for technical assistance; Andrea Patrignani and Ulrich Wagner for help in the analysis of microarray data; Maria Sofia Cattaruzza for statistical support; and Marian E. Kent for editorial assistance.

Supplemental Data

Defective DNA mismatch repair determines a characteristic transcriptional profile in proximal colon cancers

Massimiliano di Pietro, Jacob Sabates Bellver, Mirco Menigatti, Fridolin Bannwart, Annelies Schnider, Anna Russell, Kaspar Truninger, Josef Jiricny and Giancarlo Marra.

Supplemental Information on Experimental Procedures
2 Supplemental Tables
1 Supplemental Figure

Supplemental Information on Experimental Procedures

Microarray analysis

Raw gene-expression data generated by MAS software (Affymetrix) were imported for analysis into the GeneSpring software program (Silicon Genetics) and subjected to a double normalization procedure: per chip (normalization to the 50th percentile of all values on a given array) and per gene (normalization to the median expression level of the given gene across all samples). Data interpretation was carried out in the log-of-ratio mode with the cross-gene error model turned on. Genes with low expression levels were excluded from analysis by filtering out probes with "absent calls" and/or raw signals < 200 in more than 50% of the samples in each group.

Unsupervised and supervised hierarchical clustering analyses were performed using the Pearson correlation coefficient to measure similarities and a separation ratio of 1 for merging similar branches. A second unsupervised approach known as Principal Component Analysis (PCA) was also used. PCA reduces the dimensionality (number of variables) of a data set while retaining most of its original variability.

It identifies independent dimensions or "components" within a multidimensional gene data set, which account for or explain variance within the set. These components are then ranked (component 1, component 2, etc.) based on the decreasing fraction of data set variance they explain. In addition, within each component, each gene is assigned a PCA value ranging from - 1.0 to 1.0, which indicates its relative impact on that component. Values at the extremes of this range indicate maximal influence, while those closer to zero reflect minimal influence. For our analysis, the GeneSpring program identified the nine dimensions (principal components 1-9) that accounted for the highest percentages of global variability in the data set. We focused our attention on components 1 and 2, i.e., those explaining the largest amounts of variance, and attempted to correlate them with the MMR phenotype of cancers. For the purposes of our analysis, only PCA values between 0.7 and 1.0 (or -0.7 and -1.0) were considered to be significant.

For supervised statistical group comparison analyses, a non-parametric test (Mann-Whitney) was used with a False Discovery Rate of 0.05 and Benjamini-Hochberg Multiple Testing correction; for comparison of normal mucosal tissues and cancers, which yielded larger inter-group differences, Bonferroni correction was used to increase the stringency of the analysis.

Flow Cytometry

The expression of the 4-1BBL protein was evaluated in colon-cancer cell lines by flow cytometry. After 5 minutes of incubation in 20 mM EDTA in PBS, 1.2×10^6 cells from each line were harvested, washed with PBS, and incubated for 30 minutes at $4^{\circ}C$ in blocking buffer (1% BSA in PBS). After centrifugation, the cell pellet was

resuspended in FACS buffer (2% FCS and 0.01% sodium azide in PBS) containing 500 µg/ml of the monoclonal antibody against 4-1BBL (anti-human CD137, Immatics) and incubated for 2 hours on ice. A control incubation with isotype IgG1 was performed for each sample. Cells were then washed twice in FACS buffer and incubated in this buffer with the secondary anti-mouse goat FITC-conjugated antibody for 45 minutes on ice. After two additional washings in FACS buffer, cells were fixed in 1% formaldehyde in PBS and analyzed in a Coulter Epics Altra flow cytometer (Beckman-Coulter). Data were analyzed using Microsoft WinMDI Expo 32 software (version 1.28).

Supplemental Table S1. Genes displaying significantly§ different expression levels in proximal colon cancers compared with normal mucosal specimens (regardless of MMR status).

Category	Affymetrix ID	GeneBank	Name	Description	Fold differences #
Genes down-regulated in cancers compared to normal mucosa					
Cytosol	38848_at	AI732905	*ZG16*	zymogen granule protein	322
	40739_at	M83670	*CA4*	carbonic anhydrase IV	36
	40095_at	J03037	*CA2**	carbonic anhydrase II	27
	34572_at	M76424	*CA7*	carbonic anhydrase VII	11
	40808_at	U03749	*CGA**	chromogranin A	23
	37478_at	Y16752	*SCGN*	secretagogin	18
	926_at	J03910	*MT1G*	metallothionine 1G	6
	32485_at	X00371	*MB*	myoglobin	4
Structural	40694_at	X73502	*KRT20**	cytokeratin 20	10
	38586_at	M10050	*L-FABP**	liver fatty acid binding protein	8
	38587_at	M18079	*I-FABP*	intestinal fatty acid binding protein	7
	38800_at	D45352	*STMN2*	stathmin-like 2	6
	35974_at	U10485	*JAW1*	lymphoid-restricted membrane protein	6
	34595_at	AF105424	*MYO1A*	brush border myosin I	5
	39055_at	M32886	*SCN*	calcium-binding protein; sorcin	5
Metabolism	33672_f_at	U59209	*UGT2B17*	UDP glycosyltransferase 2 B17	36
	33068_f_at	U08854	*UGT2B15*	UDP glycosyltransferase 2 B15	32
	36247_f_at	M12272	*ADH3*	alcohol dehydrogenase 1 gamma subunit	33
	34637_f_at	M12963	*ADH1*	alcohol dehydrogenase 1 alpha subunit	27
	34002_at	M77144	*HSD3B2*	hydroxy-delta-5-steroid dehydrogenase	20
	38178_at	L40802	*HSD17B2*	17-beta-hydroxysteroid dehydrogenase	16
	32392_s_at	M57951	*UGT2*	bilirubin UDP-glucuronosyltransferase isozyme 2	15
	39385_at	M22324	*ANPEP**	aminopeptidase N	12
	35730_at	X03350	*ADH2*	alcohol dehydrogenase 1 beta subunit	8
	39697_at	U26726	*HSD11K*	hydroxysteroid (11-beta) dehydrogenase 2	9
	40882_at	Y09616	*CE-2 ICE*	carboxylesterase 2	5
	32570_at	L76465	*PGDH1*	hydroxyprostaglandin dehydrogenase 15-(NAD)	5
	33382_at	M92449	*ASAHL*	N-acylsphingosine amidohydrolase	5
	33702_f_at	L05144	*PKC1*	phosphoenolpyruvate carboxykinase	5
	31744_at	D89479	*ST1B2*	thyroid hormone sfotransferase	5
	41770_at	AA420624	*MAOA**	monoamine oxidase A	5
	34430_at	U70732	*GPT*	glutamic-pyruvate transaminase	4
	39408_at	Z80345	*ACADS*	acyl-Coenzyme A dehydrogenase	4
	37482_at	U37100	*AKR1B10*	aldo-keto reductase	12
	40743_at	M60092	*AMPD1*	adenosine monophosphate deaminase 1	8
	39950_at	Y08136	*SMPDL3A*	sphingomyelin phosphodiesterase	5
Extracellular matrix	36569_at	X64559	*TNA**	tetranectin (plasminogen-binding protein)	4
	34363_at	Z11793	*SEPP1*	selenoprotein P	5
	35912_at	AJ010901	*MUC4*	mucin 4	5
Cytokines/secreted	39726_at	J04040	*GCG*	Glucagon	76
	33995_at	M97496	*GUCA2A**	Guanylin	71
	33512_at	Z70295	*GUCA2B**	Uroguanylin	27
	37782_at	AI636761	*SST*	somatostatin	23
	36449_s_at	D13897	*PYY**	peptide YY	17
	160040_at	X52001	*EDN3*	endothelin 3	11
	34974_at	Y13323	*ADAMDEC1*	disintegrin protease	10
	33052_at	U95301	*PLA2G10*	calcium-dependent group X phospholipase A2	5
Immune response	37006_at	AI660656	*IGJ*	immunoglobulin J	8
	37454_at	AJ001634	*CCL13*	monocyte chemotactic protein	4
	41353_at	Z29574	*TNFRSF17*	B-cell maturation factor	8
	41164_at	X67301	*IGHM*	gM heavy chain constant region	5
	39014_at	D84239	*FCGBP*	IgG Fc binding protein	13
	31315_at	D84143	*mAb59*	immunoglobulin light chain V-J region	5
	35566_f_at	AF015128	*Vh26*	IgG heavy chain variable region	4
Integral to membrane	38884_at	AF039400	*CLCA1**	chloride channel	36
	33651_at	AB013456	*AQP8**	aquaporin 8	42
	35285_at	AF007216	*SLC4A4*	sodium bicarbonate cotransporter	27
	39637_at	U14528	*SLC26A2*	sulfate anion transporter 1	21
	928_at	L02785	*SLC26A3**	down-regulated in adenoma protein	15
	35525_at	AB020527	*SLC17A4*	Na/PO4 cotransporter	10
	160035_at	X98311	*CEACAM7**	carcinoembryonic antigen	24
	36082_at	S71326	*CEACAM1**	carcinoembryonic antigen	6
	33007_at	AC002302	*LOC63928*	hepatocellular carcinoma antigen	21
	33610_at	AL049977	*CLDN8*	claudin 8	15

	35717_at	AB020629	*ABCA8*	ATP-binding cassette, sub-family A member 8	10
	33733_at	AF093771	*ABCG2*	ATP-binding cassette, sub-family G member 2	8
	39682_at	X87159	*SCNN1B*	sodium channel, nonvoltage-gated 1, beta	10
	31976_at	X81333	*MEP1B*	meprin A, beta	9
	36275_at	AB002438	*SEMA6A*	semaphorin 6A	7
	39087_at	U28249	*FXYD3*	ion transport	7
	31645_at	AB020625	*BTNL3*	butyrophilin-like 3	6
	40666_at	AF039918	*ENTPD5*	nucleoside-diphosphatase	5
	33462_at	D13626	*GPR105*	putative G-protein-coupled receptor	4
	35449_at	U11276	*KLRB1*	killer cell lectin-like receptor subfamily B	5
	33766_at	X77777	*VIPR1*	vasoactive intestinal peptide receptor 1	4
	37875_at	U79725	*GPA33*	intestine-specific antigen	5
	36454_at	AF037335	*CA12*	carbonic anhydrase XII	4
	41158_at	M54927	*PLP1*	proteolipid protein 1	5
Transcription	38519_at	U68233	*FXR*	retinoid receptor	13
regulators	36214_at	U70663	*KLF4**	Kruppel-like factor 4	7
	33249_at	M16801	*NR3C2*	mineralocorticoid receptor	6
	36239_at	Z49194	*POU2AF1*	B cell-specific	5
Nuclear proteins	37972_at	U75744	*DNASE1L3*	deoxyribonuclease I-like 3	7
Signalling	38408_at	L10373	*TM4SF2*	transmembrane 4 superfamily member 2	8
	1970_s_at	Z71929	*FGFR2*	fibroblast growth factor receptor 2	6
	40330_at	AL050031	*PLCCE1*	phospholipase C epsilon 1	5
	31782_at	U31099	*PTGDR*	prostaglandin D2 receptor (DP)	5
	36215_at	M54181	*PRKACB*	cAMP-dependent protein kinase	5
	32521_at	AF056087	*SFRP1*	secreted frizzled-related protein 1	4
	40367_at	M22489	*BMP2*	bone morphogenetic protein 2A	4
	33709_at	AF067224	*PDE9A2*	cGMP phosphodiesterase A2	5
Unknown function	33613_at	AA806239	*EST*		12
	37821_at	AF041260	*BCAS1**	breast carcinoma amplified sequence 1	8
	35674_at	AB023211	*PADI2*	peptidyl arginine deiminase	7
	36066_at	AB020635	*KIAA0828*	putative adenosylhomocysteinase 3	7
	32735_at	AB023148	*KIAA0931*		6
	34043_at	AL022724	*C6orf105*	chromosome 6 open reading frame 105	6
	33500_i_at	S71043	*MGC27165*	hypothetical protein	5
	32110_at	AB011095	*KIAA0523*	sulfotransferase activity	5
	35226_at	U71207	*EYA2*	possible transcriprion co-activator	5
	37492_at	AB007969	*KIAA0500*		4

Genes up-regulated in cancers compared to normal mucosa

Extracellular matrix	37892_at	J04177	*COL6**	alpha-1 (type XI) collagen precursor	80
	35474_s_at	Y15915	*COL1A1*	collagen alpha 1 type I	20
	32305_at	J03464	*COL1A2*	pre-pro-alpha-2 type I collagen	13
	38420_at	Y14690	*COL5A2**	collagen, type V, alpha 2	6
	39333_at	M26576	*COL4A1*	collagen, type IV, alpha 1	5
	37459_at	X57527	*COL8A1*	collagen, type VIII, alpha 1	4
	32488_at	X14420	*COL3A1*	collagen, type III, alpha 1	4
	1693_s_at	D11139	*TIMP 1**	tissue inhibitor of metalloproteinase 1	10
	668_s_at	L22524	*MMP-7*	matrilysin (metalloproteinase)	72
	38428_at	M13509	*MMP-1**	matrix metalloproteinase 1	25
	437_at	X05232	*STMY1**	matrix metalloproteinase 3	20
	38181_at	X57766	*STMY3*	matrix metalloproteinase 11	5
	31859_at	J05070	*CLG4B*	matrix metalloproteinase 9	6
	38125_at	M14083	*PAI-1**	plasminogen activator inhibitor	10
	37310_at	X02419	*uPA**	urokinase-plasminogen activator	5
	38112_g_at	X15998	*CSPG2*	chondroitin sulfate proteoglycan 2	8
	33127_at	U89942	*WS9-14*	lysyl oxidase-like 2	7
	32818_at	X78565	*TNC**	tenascin C	6
	35832_at	AB029000	*Sulf1**	Sulphatase 1	27
	32314_g_at	M12125	*TPM2**	fibroblast tropomyosin 2	4
Cytokines	1369_s_at	X78565	*IL-8**	interleukin 8	173
	38299_at	X04430	*IL-6**	Inteleukin 6 (interferon-beta-2)	33
	1520_s_at	M15330	*IL1B**	Interleukin 1-beta	20
	37603_at	X52015	*IL1RA*	interleukin-1 receptor antagonist.	14
	1358_s_at	U22970	*gene 6-16*	interferon-inducible peptide precursor	7
	658_at	L12350	*THBS2*	thrombospondin 2	21
	36197_at	Y08374	*HCGP-3P*	chitinase 3-like 1	21
	37187_at	M36820	*GRO-beta*	GRO-2 oncogene	12
	408_at	X54489	*MSGA*	GRO-1 oncogene	10
	34022_at	M36821	*GRO-gamma*	GRO-3 oncogene	8
	31720_s_at	M10905	*FN1*	fibronectin	8
	41354_at	U25997	*STC**	stanniocalcin 1	9
	38126_at	J04599	*DSPG1*	proteoglycan I precursor; biglycan	8

	1890_at	AB000584	*MIC-1**	TGF-beta superfamily protein	8
	671_at	J03040	*SPARC**	osteonectin	5
Cytosol	41471_at	W72424	*S100A9**	Calgranulin B	19
	41096_at	AI126134	*S100A8**	Calgranulin A	8
	1069_at	U04636	*COX-2**	cyclooxygenase-2	9
	32186_at	M80244	*SLC7A5*	solute carrier family 7, member 5	7
	36979_at	M20681	*GLUT-3*	solute carrier family 2, member 3	5
	34793_s_at	M22299	*PLS3*	T-plastin	4
	39070_at	U03057	*HSN**	Fascin homolog-1	4
Immune response &	39945_at	U09278	*FAPA**	fibroblast activation protein	88
integral to membrane	160031_at	X63629	*PCAD*	placental-cadherin	10
	2087_s_at	D21254	*CAD11 OB*	osteoblast-cadherin	5
	39395_at	AA704137	*THY-1*	membrane glycoprotein	7
	41870_at	AF030428	*HT1A-1*	lung type-I cell membrane-associated prot	7
	37360_at	U66711	*TSA-1*	lymphocyte antigen 6 complex	6
	1507_s_at	D11151	*EDNRA**	endothelin-A receptor	5
	2086_s_at	D17517	*TYRO3*	protein tyrosine kinase	5
	2036_s_at	M59040	*CD44*	cell adhesion molecule	4
	39753_at	X06256	*ITGA-5*	integrin alpha 5 subunit precursor	5
Transcription	1898_at	L24203	*ATDC*	ataxia-telangiectasia group D-assoc prot	9
regulators	40126_at	Z97200	*PHOX1*	Paired Mesoderm Homeobox 1	8
	39420_at	S62138	*TLS/CHOP*	fusion protein	5
	41037_at	U63824	*TEFR-1a**	transcriptional enhancer factor-1	5
	36021_at	AL049409	*LEF-1**	lymphoid enhancer-binding protein 1	4
	36813_at	U96131	*TRIP-13**	thyroid hormone receptor interactor 13	4
	37724_at	V00568	*MYC**	c-myc oncogene	4
	40790_at	AB004066	*DEC-1*	basic helix-loop-helix protein	4
Nuclear proteins	35909_at	Z50194	*TDAG51*	pleckstrin homology-like domain A 1	21
	37282_at	AJ000186	*MAD2L1*	mitotic arrest deficient-like 1	4
	1651_at	U73379	*UbcH10*	ubiquitin carrier protein E2-C	5
	34282_at	AB010812	*NRF3*	NF-E2-related factor 3	5
	40690_at	X54942	*CKSHS2*	CDC28 protein kinase 2	4
Signalling	661_at	L13698	*GAS1*	growth arrest-specific 1	7
	38326_at	M69199	*G0S2*	G0/G1 switch gene	8
	38750_at	U97669	*NOTCH3**	Notch (Drosophila) homolog 3	5
	39827_at	AA522530	*RTP801*	DANN-damage inducible transcript 4	4
	35692_at	AL080235	*RIS-1*	Ras-induced senescence 1	4
	1347_at	S78187	*CDC25B**	cell division cycle 25B	4
	1396_at	L27560	*IGFBP5*	insulin-like growth factor binding prot 5	4
	1586_at	M35878	*IGFBP3**	insulin-like growth factor binding prot 3	4
	40733_f_at	D89377	*MSX 2*	msh homeo box homolog 2	4
Metabolism	39372_at	W26480	*FADS1*	fatty acid desaturase 1	6
	576_at	M93718	*NOS3**	nitric oxide synthase 3	6
	37032_at	U08021	*NNMT*	nicotinamide N-methyltransferase	5
	35934_at	L19161	*eIF-2 gamma*	eukaryotic translation initiation factor 2	4
Unknown function	36070_at	AL049389	*KIAA1199*	IR2155535	28
	36007_at	AL050137	*DKFZp586L151*	hypothetical protein	6
	1842_at	HG2724	*Tls/Chop*	fusion activated oncogene	5

§ Mann-Whitney test with a False Discovery Rate of 0.05 and Bonferroni correction

Minimal fold-change: 4.0 - Alterations with fold changes < 4 are not listed, even if statistically significant

Underlined: Genes displaying the same type of alteration (up- or down-regulated *versus* normal mucosa) in colon cancers and colon-cancer cell lines

* Gene expression alterations already known to be associated with colon cancer

Supplemental Table S2. Genes found differentially expressed in MMR(+) and MMR(–) colon-cancer cell lines

Genes more highly expressed in MMR(+) colon-cancer cell lines *

Category	Affymetrix ID	GeneBank	Gene Name	Gene description	Fold differences * MMR(+) *vs.* MMR(-)
Metabolism	40215_at	D50840	*UGCG*	glucosyl ceramide synthase	2.7
	37373_at	U27460	*UGP2*	UDP-glucose pyrophosphorylase 2	2.3
Structural	39754_at	X53002	*ITGB5*	integrin beta-5	4.7
Signalling	37679_at	Y10313	*IFRD1*	IFN-related developmental regulator 1	3.8
	41742_s_at	AF061034	*OPTN*	optineurin	3.1
	1814_at	D50683	*TGFbetaRII*	TGF beta Receptor II	2.6
	34268_at	X91809	*RGS19*	regulator of G-protein signalling 19	2.5
	1603_g_at	L33881	*PRKCI*	protein kinase C, iota	2.2
	1926_at	U48801	*VEGFB*	vascular endothelial growth factor B	2.2
	36083_at	U01160	*SAS*	sarcoma amplified sequence	2.2
	37003_at	X62654	*CD63*	CD63 antigen (melanoma 1 antigen)	2.1
	37045_at	D87443	*SNX19*	sorting nexin 19	2
	34852_g_at	AF011468	*aurora-A*	serine/threonine kinase 6	2
Apoptosis	37028_at	U83981	*GADD34*	protein phosphatase 1	2.7
	32260_at	X86809	*PEA15*	astrocytic phosphoprotein	2.6
	31536_at	AB020693	*RTN4/nogo*	reticulon 4	2.1
DNA repair	32870_g_at	AF073362	*MRE11A*	meiotic recombination 11 homolog A	3.3
	34684_at	L36140	*RECQL*	DNA helicase Q1-like	2.5
Transcription	34396_at	AB023195	*ASXL1*	additional sex combs like 1	2.7
	35413_s_at	M77172	*ZNF22*	zinc finger protein 22	2.5
	34195_at	AL121073	*ODAG*	ocular development-associated	2.4
	229_at	M37197	*CEBPZ*	CCAAT-box-binding factor	2.1
Nuclear	33845_at	W28483	*HNRPH1*	heterogeneous ribonucleoprotein H1	2.5
	36125_s_at	L38696	*RALY*	RNA binding protein	2.4
	35221_at	X91648	*PURA*	purine-rich binding protein A	2.4
	31873_at	U52112	*ARD1*	N-terminal acetyltransferase	2.1
Others	379_at	AB006679	*TXNDC9*	ATP binding protein	5.4
	40725_at	AF047438	*GOSR1*	Golgi SNAP receptor complex 1	2.2
	40485_at	AA176780	*TRIM44*	zinc ion binding protein	2.2
	41757_at	W25933	*VAPB*	vesicle-associated membrane protein	2.2
	39005_s_at	AB018257	*ZNF294*	zinc finger protein 294	2.1
Unknown function	34246_at	AA418437	*C6orf145*	chrom 6 open reading frame 145	10
	40517_at	AB002370	*KIAA0372*	KIAA0372 protein	3.8
	38599_s_at	AD001530	*XAP5*	putative protein	2.1
	34880_at	AC002115	*MGC10433*	hypothetical 36.5 kDa protein	2.1

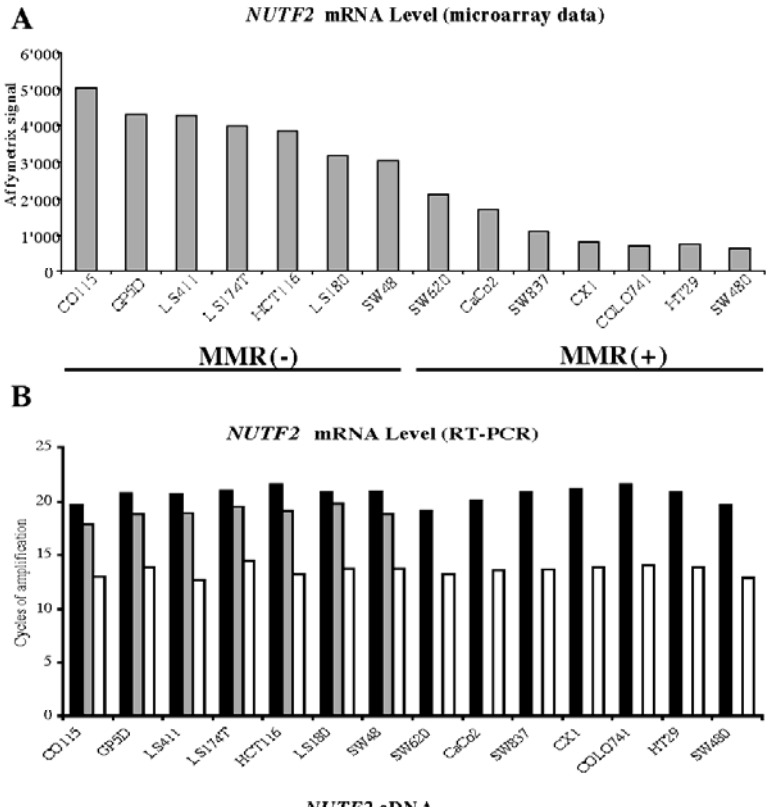
Genes more highly expressed in MMR(-) colon-cancer cell lines *

Category	Affymetrix ID	GeneBank	Gene Name	Gene description	Fold differences * MMR(-) *vs.* MMR(+)
Nuclear	35201_at	X16135	*HNRPL*	heterogeneous nuclear ribonucleoprot L	6.5
	35553_at	AF042181	*TSPYL1*	testis-specific Y-encoded-like protein	4.4
	31858_at	X07315	*NUTF2*	nuclear transport factor 2	3.4
	40348_s_at	W25866	*ANP32E*	acidic nuclear phosphoprotein 32 E	2.2
Signalling	38489_at	M60047	*FGFBP1*	FGF binding protein 1	11.1
	2031_s_at	U03106	*p21*	cyclin-dependent kinase inhibitor 1A	6.7
	40454_at	X87241	*FAT*	FAT tumor suppressor homolog 1	3.5
	509_at	U44378	*SMAD4*	SMAD, mothers against DPP hom 4	3.2
	32445_at	D63390	*PAFAH1B2*	platelet-activ. factor acetylhydrolase	3.2
	36324_at	X68487	*ADORA2B*	adenosine A2b receptor	2.9
	898_s_at	L37360	*EFNA3*	ephrin-A3	2.6
	40777_at	X87838	*CTNNB1*	beta-catenin	2.6
	38013_at	AL096842	*MTUS1*	mitochondrial tumor suppressor 1	2.4
Apoptosis	198_g_at	U29656	*nm23*	non-metastatic cells 3	3.6
	34449_at	U13022	*Caspase 2*	apoptosis-related protease	3.4
	38871_at	AJ006288	*BCL10*	B-cell CLL/lymphoma 10	2.9
Immune response	35000_at	U03398	*4-1 BBL*	TNF superfamily member 9	6
	37844_at	AI263885	*WSX-1*	interleukin 27 receptor, alpha	2.6
Metabolism	36475_at	Z97630	*GCAT*	glycine C-acetyltransferase	3
	38503_at	M63967	*ALDH1B1*	aldehyde dehydrogenase B1	3
	36599_at	M55905	*ME2*	malic enzyme 2	2.7
	34255_at	AF059202	*DGAT1*	diacylglycerol O-acyltransferase 1	2.3
Extracellular	862_at	U04313	*serpin B5*	serine proteinase inhibitor	6.1
Adhesion	33468_at	Z26317	*DSG 2*	desmoglein 2	2.6
Transcription	37043_at	AL021154	*ID3*	Inhibitor of DNA binding 3	6.2
Others	40647_at	Z32684	*XK*	Kell blood group precursor	3.3
	40354_at	AB023421	*APG-1*	heat shock protein (hsp110 family)	2.4
Unknown	32527_at	AI381790	*APM2*	chrom 10 open reading frame 116	5.6
	36821_at	AL050367	*LOC221061*	chrom 10 open reading frame 38	4
	38250_at	D26488	*KIAA0007*	KIAA0007 protein	3.1
	33844_at	AA160724	*RCD1h*	similar to RCD1 homolog 1	2.3

* Statistical significance was assessed with the Mann-Whitney test with a False Discovery Rate of 0.05 and Benjamini-Hochberg Multiple Testing correction. Minimal fold difference: 2.0; Genes whose differential expression was characterized by fold differences < 2.0 are not shown, even if statistically significant.

Underlined: Genes displaying similar patterns of differential expression [MMR(+) *vs* . MMR(-)] in both colon cancers and colon-cancer cell lines (No fold change restriction was applied to colon cancers)

Supplemental Figure S1, *di Pietro et. al.*



Supplemental Figure S1 - *NUTF2* mRNA expression in MMR(-) and MMR(+) colon-cancer cell lines. A. Gray bars represent raw signals for *NUTF2* expression detected in MMR(-) (*x*-axis - left) and MMR(+) (*x*-axis - right) cell lines with the Affymetrix U95v2 microarray. **B.** Quantitative RT-PCR amplification of the coding region (black bars) and 3'UTR (gray bars) of *NUTF2* mRNA and the control housekeeping gene *GAPDH* (white bars) in MMR(-) (left) and MMR(+) (right) cell lines. The height of the bars, which corresponds to the number of cycles (*y*-axis) at the beginning of the log-phase amplification, is inversely correlated with the abundance of transcript. No bars are shown for the amplification reaction of the 3'UTR in MMR(+) cell lines because no PCR product was obtained by RT-PCR. The RT-PCR findings were consistent with the results of the microarray analysis because, for this gene (and most of the others represented on the array), the Affymetrix chip includes probes complementary to both the 3'UTR and the 3' end of the coding region. **C.** The poly(T) repeat in the 3'UTR of *NUTF2* cDNA is shown in red, the coding sequence in blue. Primers for the amplification reactions shown in Panel B are indicated by bold-faced type.

6.4. Appendix IV

“The transcriptome of ileal and colonic normal mucosa” Jacob Sabates-Bellver, Tobias Gonzenbach, Endre Laczko, Hubert Rehrauer, Caroline Maake, Fridolin Bannwart, Annelies Schnider, Josef Jiricny and Giancarlo Marra. Manuscript in preparation

The transcriptome of ileal and colonic normal mucosa

Short Title: Transcriptional differences between ileum and colon tissues

Jacob Sabates-Bellver[1], Tobias Gonzenbach[1], Endre Laczko[2], Hubert Rehrauer[2], Caroline Maake[3], Fridolin Bannwart[4], Annelies Schnider[5], Josef Jiricny[1] and Giancarlo Marra[1]*.

[1]Institute of Molecular Cancer Research, [2]Functional Genomic Center, and [3]Institute of Physiology, University of Zurich, Zurich, Switzerland.
[4]Institute of Pathology and [5]Department of Surgery, Triemli Hospital, Zurich, Switzerland.

*Corresponding author:
Giancarlo Marra, M.D., Ph.D.
Institute of Molecular Cancer Research
University of Zurich
Winterthurerstrasse 190
8057 Zurich
Switzerland
Tel. 0041 044 635 3472
Fax. 0041 044 635 3484
marra@imcr.unizh.ch

Abstract

Background & Aims. The incidence of adenocarcinomas in the small intestine, particularly in the ileum, is much lower than in the colon, which arises in the third most frequent cancer type both in men and women.

Even though the anatomical and physiological ileum/colon transition has been well described, the molecular explanations for the different cancer frequency in these two similar tissues have not been completely explained.

Our aim was to analyze the transcriptomes of colon and ileum normal mucosas to find out the pathways, and the key genes of those pathways, that could explain this phenomenon.

Methods. A prospective study was conducted on 12 paired colon and ileum normal mucosa samples collected endoscopically.

The transcriptome of the entire human genome was investigated using a well-standardized gene-expression platform.

Results. Unsupervised analyses of microarray data clearly distinguished ileum from colon normal mucosa samples. We found many conspicuous expression changes of genes involved in cell differentiation/proliferation, transcription regulators, Wnt-signaling pathway and apoptosis in these two tissues.

We focused our attention on two pro-apoptotic genes members of the cell death-inducing DFFA-like effectors family, *CIDEB* and *CIDEC*, which were very down-regulated in colon compared to the counterpart ileum samples. We also found their mRNA levels further down-regulated in colon adenomas and adenocarcinomas. Work at the protein level is currently in progress..

Conclusions. We have reported for the first time a comparative analysis of the gene expression profile of human colon and ileum normal mucosa. Among the up-regulated genes in ileum, we pointed out *CIDEB* and *CIDEC*, known to be involved in apoptosis, as two genes that may play an important role in prevention of colorectal tumorigenesis.

Introduction

The small intestine (duodenum, jejunum and ileum) constitutes about the 75% of the total length of the gastrointestinal tract (GI) but only 1-3% of the GI malignances appear in this portion (>50 times less frequent than malignances in the colon)[1]. Only around 40% of these malignances are adenocarcinomas (i.e., derived from epithelial cells) and 80% of them arise in the duodenum and jejunum (proximal small intestine) while they are rare in the longer and distal part (closer to the colon), the ileum[1]. On the opposite side, colorectal adenocarcinoma is the third most frequent cancer type both in men and women and its frequency is about 25 times higher than in the small intestine[2].

Although the anatomical and physiological transition between ileum and colon is well characterized, the molecular features underlining the different adenocarcinoma incidence in ileum and colon are not known although many studies about the effect of luminal factors[3-5] (for example the gut microflora, carcinogens, etc) and intrinsic differences[6-8] (for instance cell turnover rates, the redox status, etc...) have been complete to date. It is generally accepted that the amount and distribution of mucosal and luminal factors is different in these two GI tracts and that this might be important in modulating the predisposition to adenocarcinoma.

On the other hand, our knowledge about the molecular basis of colon adenocarcinomas has been substantially increased in the last 20 years. The sequence of events required to transform normal human colonic epithelium into sporadic adenomas and carcinomas has been documented[9-13].

Our aim in this study was to show for the first time a transcriptional analysis between the colon and the ileum normal mucosas. We think this may be the first step to find out the molecular features responsible for the different adenocarcinoma incidence in these two physic- and physiological similar tissues.

The wider vision of the molecular changes occurring in these two healthy tissues upcoming from this analysis can help us, and other scientists, to focus our effort in certain molecules or pathways which could enclose the explanations to this scenario.

Among all the genes found differently expressed in colon compared to ileum normal mucosa, those related to apoptosis, cell proliferation and differentiation, regulators of transcription, immune response and cell defense, may possibly be the most interesting ones to work with.

We have focused our attention on apoptosis and, particularly, on *CIDEB* and *CIDEC* genes.

Apoptosis is the regulated, or programmed, cell death of those cells that are not necessary anymore or that are potentially dangerous for the organism[14]. A family of cysteine proteases, called caspases, have been shown to play a key role in this process[15, 16]. Another important group of pro- and anti-apoptotic regulators are the Bcl-2 family, in which the protein-protein interactions are very important to understand their mechanism of action[17-19]. Although many high-quality articles have been published up to date, there's still a lot of work to do on the understanding of the regulation of this process and on the development of new therapies for human diseases[20].

CIDEB and *CIDEC*, also called cell death-inducing DFFA-like effector b and c, respectively, are members of the CIDE family of genes (*CIDEA*, *CIDEB* and *CIDEC*). These genes have homology to the N-terminal region of *DFFA* (*DFF40/CAD*) and *DFFB* (*DFF45/CAD*) (DNA fragmentation factor, 45kDa, alpha polypeptide and 40kDa, beta polypeptide, respectively)[21]. DNA fragmentation factor (DFF) is a heterodimeric protein of 40-kD (DFFB) and 45-kD (DFFA) subunits. DFFA is the substrate for caspase-3 and triggers DNA fragmentation during apoptosis. DFF becomes activated when DFFA is cleaved by caspase-3 and dissociate from DFFB, its inhibitor. DFFA has been found to trigger both DNA fragmentation and chromatin condensation during apoptosis[22, 23].

It has also been shown that as well as the caspase-3 triggered activation of DFFA, the nuclease activity of this gene seems to be regulated by the interaction between DFFB and CIDEB. CIDEB, thanks to the strong homology to the regulatory CIDE-N domain of DFFA and DFFB and to the fact that CIDE-N/CIDE-N interaction between DFFB and CIDEB is stronger than between DFFA and CIDEB, sequesters DFFB from the heterodimer complex that forms with DFFA, triggering apoptosis[24].

Since has been described that CIDEC most probably have an important role in prevention of tumorigenesis[25] and based on the strongly cell-specific hyper- or hipo-metylation dependent regulation of *CIDEB*'s promoter[26], we assume that CIDEB and CIDEC may play an important role in colorectal tumorigenesis prevention.

Material and Methods

Colon and Ileum normal mucosa samples

Samples from right-side colon and from ileum were prospectively collected by F.B. and A.S. at the gastroenterology unit of the Triemli Hospital (Zurich, Switzerland) with full institutional review-board approval. The fact that the sampling was done only by two operators, using exactly the same procedure, minimized artifacts due to sampling differences. After resection, tissues were immediately immersed in RNAlater (Ambion, Huntington, UK), to preserve RNA from degradation, for subsequent microarray analysis.

Total RNA was extracted (RNeasy Mini kit, Qiagen, Hilden, Germany) from homogenized 5- to 15-mg tissue samples and its integrity was verified by capillary gel electrophoresis (Bio Analyzer, Agilent Technologies, Santa Clara, CA). Total RNA was converted to cDNA using a cDNA synthesis kit (Invitrogen, Carlsbad, CA). Double-stranded cDNA was then converted to biotin-labeled cRNA by a T7 RNA polymerase-catalyzed reaction (MEGA Script in vitro transcription kit, Ambion, Austin, TX) with biotin-containing nucleotides from LOXO (Dossenheim, Germany). Labeled cRNA were then purified (RNeasy kit, Qiagen, Hilden, Germany) and fragmented. 15 μg cRNA were hybridized with Affymetrix U133A arrays (High Wycombe, UK), which contains 22823 *in situ* synthesized oligonucleotides representing more than 14,500 genes.

Microarray Analysis

Raw gene-expression data generated by GCOS software (Affymetrix) were imported into the GeneSpring software program (Agilent) and normalized per chip (i.e., to the median of all values on a given array) and per gene (i.e., to the median expression level of the given gene across all samples). Analysis was performed using the log expression values with GeneSpring's cross-gene error model turned on. Only probes listed as "present or marginal calls" and/or with expression values ≥100 in at least 50% of the samples (6 out of 12) in either tissue groups (that means: only in colon or only in ileum, or in both groups) were included in further analyses.

Expression data were subjected to four different unsupervised analyses: 1) Hierarchical clustering using the Pearson correlation coefficient as a similarity measure and the average linkage algorithm for branch merging; 2) Principal Component Analysis (PCA) which reduces the dimensionality (number of variables) of a data set while retaining most of its variance. PCA identifies independent "components" within a multidimensional gene data set, which account for or explain variance within the set. These components are then in decreasing order based on the fraction of data set variance they explain. In addition, within each component, each gene is assigned a PCA value ranging from - 1.0 to 1.0, which indicates its relative impact on that component. Values at the extremes of this range indicate maximal influence, while those closer to zero reflect minimal influence; 3) Correlation Analysis which involved computation of Pearson correlation coefficients for all possible sample pairs and visualization of correlation values as tile plots; 4) Correspondence Analysis (CA), another dimension-reducing method[27], that was used to identify samples associated with particular gene expression levels. Typically, CA treats a matrix of *n* gene expression levels from *p* samples as a two-way contingency table (genes by samples or vice versa) with *n* and *p* specifications for the "factors" gene and sample, respectively. Hereby, each intensity value is viewed as an abundance value of a particular transcript in a particular sample. Like PCA, CA identifies independent "factorial components" within a multidimensional gene data set, which account for variance within the set, but in this case the components are identified and ranked according to the correlation between gene and sample scores. CA and its plot, was computed using R and the contributed packages "ade4" as well as "made4", all available from Bioconductor (http://www.bioconductor.org.)

We then selected the genes differentially expressed in colon normal mucosa compared to the ileum normal mucosa using the Mann-Whitney test, a False Discovery Rate of 0.01 and the Benjamini-Hochberg Multiple Testing correction. We then analyzed these genes to identify any biological processes from the Gene Ontology database[28] that were over-represented. ErmineJ software[29] was used for this latter analysis.

Real-Time RT-PCR

One-step real-time retro-transcription poly chain reaction (real time RT-PCR) was performed with the Roche LightCycler System using the Light Cycle-RNA Master Sybr Green I Kit (Roche, Basel, Switzerland) according to the manufacturer's instructions, 0.3 µM of each oligonucleotide primer (Microsynth, Balgach, Switzerland) and 150 ng of total RNA in 20 µl of reaction volume. Primer sequences

and RT-PCR reaction conditions are available on request. The Ct is the cycle corresponding to the beginning of the log-phase amplification and one cycle difference in Ct corresponds, theoretically, to a 2-fold change in RNA concentration. Fold changes were obtained by normalizing to GAPDH used as internal control. All of the experiments were performed in duplicate, and the specificity of each amplification product was verified by agarose gel electrophoresis. Primers are available on request.

Immunohistochemistry

Sections (4 μm) of formalin-fixed, paraffin-embedded tissues were mounted on silanized slides, deparaffinized and rehydrated. Antigen was retrieved by heating the sections in a pressure cooker (120°C for 2 min) in 10 mM citrate-buffered solution (pH 6.0). Peroxidase-blocking reagent and goat serum (DakoCytomation, Glostrup, Denmark) were used sequentially to eliminate non-specific staining due to endogenous peroxidase activity and unspecific antibody binding, respectively. The sections were incubated at 4°C for 24 hours with a 1:1000 dilution of the affinity-purified CIDEB antibody. After washing, anti-rabbit secondary antibodies conjugated to peroxidase-labeled polymer (DakoCytomation EnVision+ kit) were applied for 30 min at RT. The peroxidase activity was then developed by incubation with 3,3'diaminobenzidine chromogen solution. Negative control experiments were carried out with CIDEB antibody that had been pre-absorbed with the immunogenic peptide; alternatively, primary or secondary antibodies were omitted altogether.

Results

Twelve matched couples of right-side colonic and ileal normal mucosa samples were collected after surgical operations from12 different patients.

Microarray data analysis for the 12 colonic-ileal normal mucosa paired samples revealed that 16245 of the 22283 probes present in the HG-U133A arrays were expressed in either one or both tissue groups. Four different unsupervised analyses of the transcriptional levels of the expressed genes effectively segregate the colonic normal mucosa group from the ileal one (**Figure1**).

One thousand six hundred fifty-two of the expressed probes showed significant expression changes in colon compared to ileum normal mucosa after supervised statistical analysis (**Figure 2A**). To identify the main different biological processes in each tissue group, we performed a Gene Ontology analysis with all the differentially expressed genes. The colonic normal mucosa samples had an up-regulation of genes mainly involved in both positive regulation of cell motility and cell migration. On the other hand, genes up-regulated in the ileum normal mucosa group were related to different aspects of metabolism, protein kinase C activation and regulation of apoptosis, among others (**Table 1**).

In order to try to uncover the most important transcriptional changes between these two tissues, we have selected those genes with a fold change ≥4, the ones that more significantly distinguished the

two groups of samples (**Figure 2B** and **Table2**). This list is meant to be a starting point for future confirmation studies. We found 213 genes significantly up-regulated (n=52) or down-regulated (n=161) in colon normal mucosa compared to their counterpart ileum samples.

Because is thought that specific pathways maintain the cellular homeostasis in the normal tissue and those controlling the normal proliferation/differentiation rates in the colon could be differently regulated in the ileum, we focused on the changes in the expression levels of genes involved in these pathways as well as of numerous transcription regulators (**Table 3**). We included those genes with fold changes between 4 and 2 because small changes of such molecules could still enclose very important information.

We have also focused our attention on the Wnt signaling pathway; it's the physiological regulator of epithelial homeostasis and has been well described in the literature that the transformation process to adenoma, and lately to carcinoma, begins in the epithelial crypts as a result from qualitative, quantitative and spatial subversion of this pathway[11, 12, 30-32]. Although the list of Wnt target genes is far to be completed, many studies have been performed[33, 34]. Thus, we have analyzed the expression levels of the members and targets of this pathway in ileum and colon (Table 3).

Given the very high rates of cell renewal in ileum and colon and the importance of apoptosis for this process, we focus on the analysis of the expression level of apoptosis-related genes in these two tissues to see whether we could find relevant differences that could give a hint on the reasons why the cancer incidence is much higher in colon than in ileum (**Table 4**). Amongst those genes more expressed in ileum, we focused our attention in *CIDEB* and *CIDEC*, the two members of the cell death-inducing DFFA-like effectors expressed in these two tissues. The third member of this family, *CIDEA*, was neither expressed in the ileum nor in the colon normal mucosa samples (data not shown). The mRNA over-expression of *CIDEB* and *CIDEC* in all 12 ileum normal mucosa was conspicuous compared to the 12 counterpart colon normal mucosa samples (7.6 and 8.3 times, respectively). Interestingly, the expression of these two genes was also found down-regulated in benign and malignant colorectal tumors previously studied by our group[35, 36] (**Figure 3A**). The microarray results for *CIDEB* were confirmed by real-time RT-PCR analysis (**Figure 3B-D**). Given that for *CIDEB* two different transcripts were described[26], we also showed that the difference in expression level between ileum and colon was in the short one whereas the long one remained unchanged in both tissues. At this point, we also checked the mRNA levels in our microarray data for SP1, SP3 and HNF4-alfa because they have been shown to regulated the two different *CIDEB* transcripts (SP1 and SP3 both and HNF4-alfa only the short one) together with hypermetilation of the up-stream promoter[26]. The mRNA expression levels of these genes were not significantly changed among the different tissue groups, ileum, colon, adenomas and carcinomas, we had (data not shown).

We are currently analyzing the protein level of CIDEB and CIDEC in mouse and human tissues with preserved architecture to identify their cellular localization (work in progress, data not shown).

Discussion

Why cancer frequency in colon is tremendously higher than that in ileum even though these two gastrointestinal segments are physically and morphologically very similar? To date, little is known about the possible molecular mechanisms behind this phenomenon. We attempted to answer this question by performing a comprehensive comparison of colon and ileum normal mucosa transcriptomes using the latest cDNA microarray technology.

The transcriptome of the colon normal mucosa, compared with that of ileum normal mucosa, shows the molecular differences existing between these two tissues, although, as expected because of the similar nature of both organs, they are less numerous than between colorectal tumors and colon[35, 36].

From the 1652 probes found significantly changed between these two tissues, we have drawn up a list of 213 genes up- (n=52) or down-regulated (n=161) in colon compared to ileum with a fold change ≥4. We think that this list can be used as a starting point for future verification studies.

We have shown a complete picture of the expression changes on genes related to cell proliferation/differentiation, transcription regulators, Wnt signaling and apoptosis that may elucidate the reasons why the adenocarcinoma incidence in ileum and colon is so different.

We also focused on two pro-apoptotic genes highly up-regulated in ileum compared to colon, *CIDEB* and *CIDEC*. The expression of these two genes, members of the cell death-inducing DFFA-like effectors family, is even more decreased in colon adenomas and adenocarcinomas.

It has been shown that *CIDEB* has two different transcripts. We have revealed that is the short one which is down-regulated in colon, and even more in adenomas and adenocarcinomas of the colon, compared to ileum. These results suggest that CIDEB and CIDEC may play an important role in colorectal tumorigenesis prevention.

For *CIDEC* has already been shown that may play an important role in prevention of tumorigenesis in kidney, liver and melanoma cell lines[25] but neither *CIDEB* nor *CIDEC* decreased expression levels were yet related to increased incidence of colorectal adenocarcinomas.

Work at the protein level is currently on progress and we expect it will bring valuable information about their cellular localization and putative roles against carcinogenesis.

Figure Legends

Figure 1. Unsupervised Analyses of microarray data

(**A**) *Hierarchical clustering analysis*: The 24 tissue samples represented on the *x*-axis include 12 colon (red branches) and 12 ileum (green branches) normal mucosal samples. Each of the 16245 expressed probes plotted on the *y*-axis is color-coded to indicate the expression level of the gene relative to its median level of expression across the entire set of tissue samples (blue: low; red: high). (**B**) *Principal*

component analysis (PCA). Profile plot of the normalized first principal component (PCA-1) across the 24 specimens (green dots: colon normal mucosa; red dots: ileum normal mucosa). The two tissue groups are significantly different in terms of the first principal component ($p < 0.0001$), which accounted for 30% of the total variance. (**C**) *Correlation Analysis*. Tile-plot visualization of the pair-wise correlations of the samples. The correlation matrix has 24 rows and columns corresponding to the 24 samples. The gray scale column at the right side of the tile-plot refers to the correlation values (white: high correlation; black: low correlation). The matrix shows that colon or ileum normal mucosa samples are highly correlated among each other (lower left and upper right part). *Colon and ileum normal mucosa samples are poorly* correlated (upper left and lower right part), however, the samples from the same patient generally show a higher correlation than colon and ileum samples from different patients (bright pixels on the secondary diagonals in the upper left and lower right part). This finding probably reflects the strong influence of several factors, including the individual genetic background and life-style. (**D**) *Correspondence Analysis (CA)* of mRNA log(intensity) values of expressed genes from the 12 tissue pairs (colon normal mucosa, red dots; ileum normal mucosa, green dots). The areas delimited by the ellipses represent 95% of the estimated bi-normal distribution of the sample scores on the first and second CA axes. The map of the sample scores on the first two axes shows that CA efficiently discriminates between colon and ileum normal mucosa samples.

Figure 2. Supervised Hierarchical Analysis of microarray data

(**A**) All *statistically significant probes*: As in Figure 1, 12 colon (red branches) and ileum (green branches) normal mucosa samples are represented in the *x*-axis. In this case, on the *y*-axis, we have plotted all the probes that passed the statistical analysis (Mann-Whitney test, False Discovery Rate of 0.01 and the Benjamini-Hochberg Multiple Testing correction). Each of the 1652 statistically significant probes plotted on the *y*-axis is color-coded to indicate the expression level of the gene relative to its median level of expression across the entire set of tissue samples (blue: low; red: high). (**B**) S*tatistically significant probes with a fold change ≥4*: As in Figure 1, 12 colon (red branches) and ileum (green branches) normal mucosa samples are represented in the *x*-axis. On the *y*-axis, only the 267 statistically significant probes with a fold change ≥4 (representing 213 different genes) are plotted. The color-code is the same as in (A).

Figure 3. CIDEB and CIDEC messenger RNA expression in tumors and normal mucosa of the colon and ileum. **(A)** *CIDEB* and *CIDEC* mRNA levels detected in samples of ileum and colon normal mucosa and colorectal adenomas and carcinomas (normalized data of the raw signal values (*y*-axis)). Whereas *CIDEB* and *CIDEC* mRNA levels in ileal normal mucosa were 7.6 and 8.3 times higher than those found in the colonic normal mucosa, levels in colorectal adenomas and carcinomas were at least 2 times lower. All these differences were highly significant (*$p < 0.001$; **$p < 2.5e\text{-}7$; ***$p < 1.2e\text{-}8$). **(B)** Real-time RT-PCR amplification of the two different transcripts of *CIDEB* and the control housekeeping gene *GAPDH* in one ileum and one colon normal mucosa and in one colorectal adenoma and one carcinoma

representative samples. A total of 5 samples for each tissue were examined in triplicate. Ct (threshold cycle): number of cycles (*x*-axis) at the beginning of the log phase of amplification. After 37 cycles of PCR, the long transcript of *CIDEB* (Lt) was equally amplified in all samples whereas the short transcript (St) was increasingly amplified in ileum, colon, adenomas and carcinoma samples. **(C)** Melting curve picks for Lt and St *CIDEB* and *GAPDH*. The specificity of the products obtained is very clear. **(D)** The *CIDEB* amplification products were subjected to agarose-gel electrophoresis to verify the specificity of the amplification. M: 100-bp ladder marker; NC: Negative Control; IL: ileum normal mucosa; CO: colon normal mucosa; Ad: adenoma; CRC: carcinoma.

References

1. De Vita V. HS, and Rosenberg SA. Cancer of the gastrointestinal tract, chapter 33 in "Cancer, Principles & Practice of Oncology". Lippincott Williams & Wilkins, 2001.
2. Jemal A, Murray T, Ward E, Samuels A, Tiwari RC, Ghafoor A, Feuer EJ, Thun MJ. Cancer statistics, 2005. CA Cancer J Clin 2005;55:10-30.
3. Guarner F. Enteric flora in health and disease. Digestion 2006;73 Suppl 1:5-12.
4. Hope ME, Hold GL, Kain R, El-Omar EM. Sporadic colorectal cancer--role of the commensal microbiota. FEMS Microbiol Lett 2005;244:1-7.
5. O'Hara AM, Shanahan F. The gut flora as a forgotten organ. EMBO Rep 2006;7:688-93.
6. Lowenfels AB. Etiological aspects of cancer of the gastrointestinal tract. Surg Gynecol Obstet 1973;137:291-9.
7. Lowenfels AB. Why are small-bowel tumours so rare? Lancet 1973;1:24-6.
8. Sanders LM, Henderson CE, Hong MY, Barhoumi R, Burghardt RC, Carroll RJ, Turner ND, Chapkin RS, Lupton JR. Pro-oxidant environment of the colon compared to the small intestine may contribute to greater cancer susceptibility. Cancer Lett 2004;208:155-61.
9. Kinzler KW, Vogelstein B. Life (and death) in a malignant tumour. Nature 1996;379:19-20.
10. Kinzler KW, Vogelstein B. Landscaping the cancer terrain. Science 1998;280:1036-7.
11. Korinek V, Barker N, Morin PJ, van Wichen D, de Weger R, Kinzler KW, Vogelstein B, Clevers H. Constitutive transcriptional activation by a beta-catenin-Tcf complex in APC-/- colon carcinoma. Science 1997;275:1784-7.
12. Morin PJ, Sparks AB, Korinek V, Barker N, Clevers H, Vogelstein B, Kinzler KW. Activation of beta-catenin-Tcf signaling in colon cancer by mutations in beta-catenin or APC. Science 1997;275:1787-90.
13. Vogelstein B, Kinzler KW. The multistep nature of cancer. Trends Genet 1993;9:138-41.
14. Hengartner MO. The biochemistry of apoptosis. Nature 2000;407:770-6.
15. Alnemri ES, Livingston DJ, Nicholson DW, Salvesen G, Thornberry NA, Wong WW, Yuan J. Human ICE/CED-3 protease nomenclature. Cell 1996;87:171.
16. Kerr JF, Wyllie AH, Currie AR. Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics. Br J Cancer 1972;26:239-57.
17. Adams JM, Cory S. The Bcl-2 protein family: arbiters of cell survival. Science 1998;281:1322-6.
18. Antonsson B, Martinou JC. The Bcl-2 protein family. Exp Cell Res 2000;256:50-7.
19. Reed JC. Double identity for proteins of the Bcl-2 family. Nature 1997;387:773-6.

20. Kaufmann SH, Hengartner MO. Programmed cell death: alive and well in the new millennium. Trends Cell Biol 2001;11:526-34.
21. Inohara N, Koseki T, Chen S, Wu X, Nunez G. CIDE, a novel family of cell death activators with homology to the 45 kDa subunit of the DNA fragmentation factor. Embo J 1998;17:2526-33.
22. Widlak P. The DFF40/CAD endonuclease and its role in apoptosis. Acta Biochim Pol 2000;47:1037-44.
23. Widlak P, Li P, Wang X, Garrard WT. Cleavage preferences of the apoptotic endonuclease DFF40 (caspase-activated DNase or nuclease) on naked DNA and chromatin substrates. J Biol Chem 2000;275:8226-32.
24. Lugovskoy AA, Zhou P, Chou JJ, McCarty JS, Li P, Wagner G. Solution structure of the CIDE-N domain of CIDE-B and a model for CIDE-N/CIDE-N interactions in the DNA fragmentation pathway of apoptosis. Cell 1999;99:747-55.
25. Liang L, Zhao M, Xu Z, Yokoyama KK, Li T. Molecular cloning and characterization of CIDE-3, a novel member of the cell-death-inducing DNA-fragmentation-factor (DFF45)-like effector family. Biochem J 2003;370:195-203.
26. Da L, Li D, Yokoyama KK, Li T, Zhao M. Dual promoters control the cell-specific expression of the human cell death-inducing DFF45-like effector B gene. Biochem J 2006;393:779-88.
27. J-P B. Correspondence Analysis Handbook. Marcel Dekker Inc., 1992.
28. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000;25:25-9.
29. Lee HK, Braynen W, Keshav K, Pavlidis P. ErmineJ: tool for functional analysis of gene expression data sets. BMC Bioinformatics 2005;6:269.
30. Gregorieff A, Pinto D, Begthel H, Destree O, Kielman M, Clevers H. Expression pattern of Wnt signaling components in the adult intestine. Gastroenterology 2005;129:626-38.
31. Pinto D, Clevers H. Wnt control of stem cells and differentiation in the intestinal epithelium. Exp Cell Res 2005;306:357-63.
32. Pinto D, Clevers H. Wnt, stem cells and cancer in the intestine. Biol Cell 2005;97:185-96.
33. van de Wetering M, Sancho E, Verweij C, de Lau W, Oving I, Hurlstone A, van der Horn K, Batlle E, Coudreuse D, Haramis AP, Tjon-Pon-Fong M, Moerer P, van den Born M, Soete G, Pals S, Eilers M, Medema R, Clevers H. The beta-catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. Cell 2002;111:241-50.
34. Van der Flier LG, Sabates-Bellver J, Oving I, Haegebarth A, De Palo M, Anti M, Van Gijn ME, Suijkerbuijk S, Van de Wetering M, Marra G, Clevers H. The Intestinal Wnt/TCF Signature. Gastroenterology 2007;132:628-32.
35. di Pietro M, Sabates Bellver J, Menigatti M, Bannwart F, Schnider A, Russell A, Truninger K, Jiricny J, Marra G. Defective DNA mismatch repair determines a characteristic transcriptional profile in proximal colon cancers. Gastroenterology 2005;129:1047-59.
36. Sabates-Bellver J, Van der Flier, L.G., de Palo, M., Cattaneo, E., Maake, C., Rehrauer, H., Laczko, E., Kurowski, M.A., Bujnicki, J.M., Menigatti, M., Luz, J., Ranalli, T.V., Pastorelli, A., Faggiani, R., Anti, M., Jiricny, J., Clevers, H., Marra, G. The transcriptome of human colorectal adenomas: Identification of KIAA1199 as a novel tumor biomarker. Submitted paper.

Figure 1. Sabates-Bellver et al.

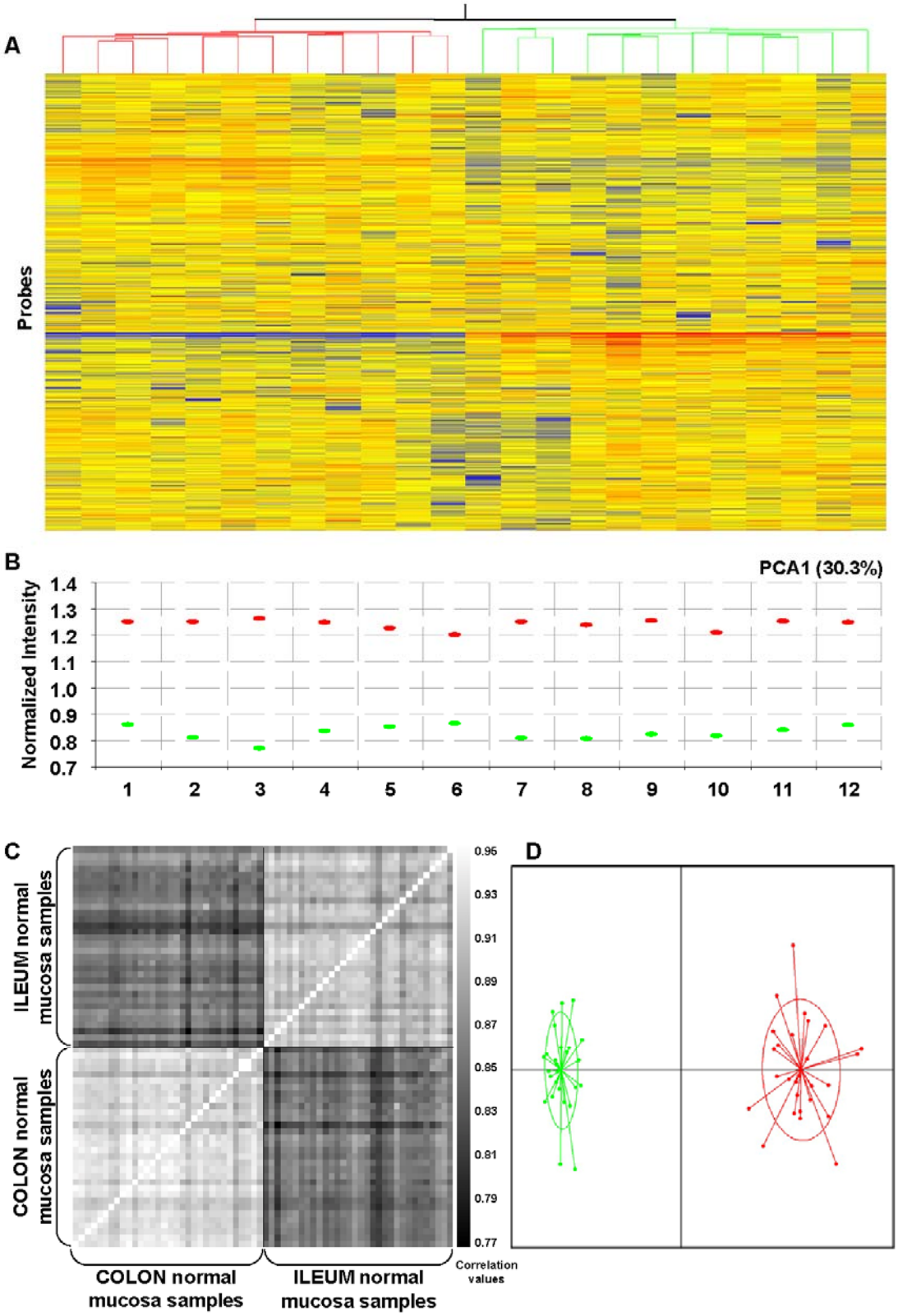


Figure 2. Sabates-Bellver et al.

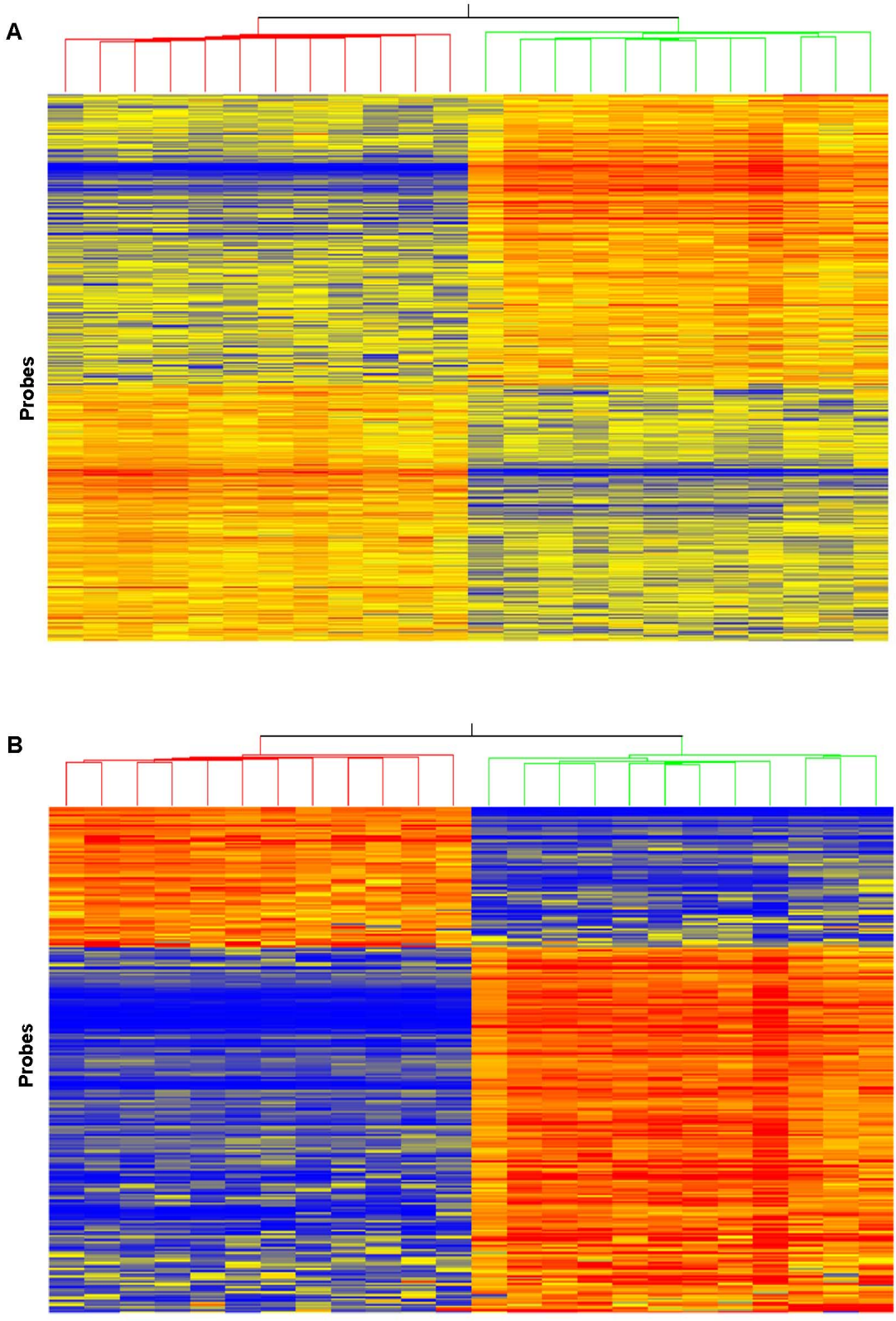


Figure 3, Sabates-Bellver et al.

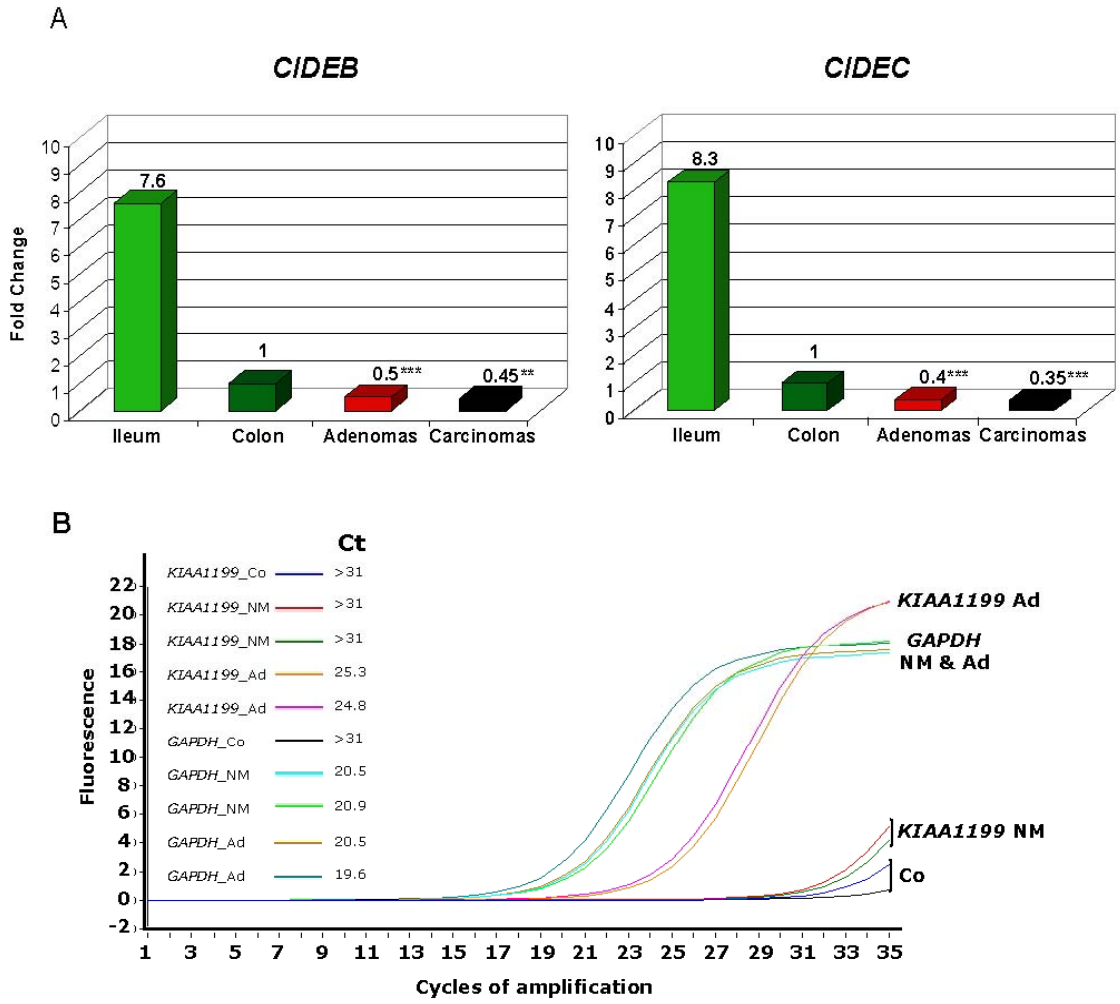


Table 1: *Biological-process* categories most significantly over-represented in the set of genes differentially expressed in colon vs. ileum normal mucosa (ErmineJ' overrepresentation analysis. Only biological processes containing between 5 and 200 genes were considered)

Genes more expressed in colon than in ileum normal mucosa

Gene Ontology category	Category genes *	Significant genes ^	p-value #
positive regulation of cell motility	7	4	2.36E-03
glycolysis	43	10	3.83E-03
positive regulation of cell migration	8	4	4.03E-03
anion transport	23	6	9.77E-03
one-carbon compound metabolism	30	7	9.86E-03

Genes more expressed in ileum than in colon normal mucosa

Gene Ontology category	Category genes *	Significant genes ^	p-value #
carboxylic acid metabolism	121	31	3.30E-11
amino acid and derivative metabolism	42	16	9.21E-09
amine metabolism	48	17	7.23E-09
cellular lipid metabolism	169	32	2.69E-08
response to xenobiotic stimulus	26	12	3.47E-08
xenobiotic metabolism	23	11	7.79E-08
steroid metabolism	93	21	4.83E-07
carbohydrate metabolism	190	31	1.21E-06
response to drug	20	8	4.23E-05
catabolism	39	11	6.03E-05
digestion	35	10	1.28E-04
hexose metabolism	42	11	1.22E-04
catecholamine metabolism	13	6	1.33E-04
circulation	75	15	1.46E-04
lipid catabolism	37	10	1.74E-04
glycerol metabolism	14	6	2.06E-04
polyol metabolism	10	5	2.56E-04
phenol metabolism	10	5	2.42E-04
sodium ion transport	62	13	2.30E-04
urea metabolism	7	4	4.44E-04
fructose metabolism	11	5	4.35E-04
aldehyde metabolism	11	5	4.15E-04
protein kinase C activation	16	6	4.35E-04
regulation of apoptosis	97	16	6.81E-04
biogenic amine metabolism	12	5	6.97E-04
urea cycle	8	4	8.70E-04
glycerol-3-phosphate metabolism	8	4	8.38E-04
detection of external stimulus	13	5	1.10E-03
nitrogen compound metabolism	13	5	1.06E-03
cellular catabolism	19	6	1.27E-03
detection of biotic stimulus	14	5	1.66E-03
metal ion transport	173	22	1.94E-03
response to chemical stimulus	162	21	1.96E-03
protein amino acid N-linked glycosylation	35	8	2.25E-03
regulation of neurotransmitter levels	15	5	2.34E-03
blood pressure regulation	15	5	2.28E-03
neurotransmitter metabolism	10	4	2.51E-03
nitric oxide metabolism	10	4	2.44E-03

detection of pest, pathogen or parasite	10	4	2.38E-03
nitric oxide biosynthesis	10	4	2.32E-03
carboxylic acid transport	44	9	2.34E-03
phospholipid biosynthesis	22	6	2.67E-03
sphingolipid catabolism	6	3	2.74E-03
regulation of immune response	6	3	2.68E-03
regulation of blood vessel size	6	3	2.62E-03
bile acid metabolism	6	3	2.56E-03
glutamine family amino acid metabolism	16	5	2.66E-03
fatty acid metabolism	104	15	3.03E-03
amine transport	38	8	3.12E-03
neurotransmitter transport	23	6	3.07E-03
icosanoid metabolism	11	4	3.19E-03
glucose transport	11	4	3.13E-03
biosynthesis	108	15	4.30E-03
hexose transport	12	4	4.93E-03
G-protein signaling, coupled to IP3 second messenger (phospholipase C activating)	42	8	6.25E-03
membrane lipid biosynthesis	19	5	6.58E-03
biological process unknown	91	13	6.52E-03
amino acid transport	34	7	6.42E-03
cholesterol metabolism	52	9	7.01E-03
glutamine family amino acid biosynthesis	13	4	6.90E-03
feeding behavior	13	4	6.78E-03
glutathione metabolism	8	3	8.13E-03
alkene metabolism	8	3	8.00E-03

* Number of genes annotated in the corresponding Gene Ontology category

^ Number of genes from each category that were found significantly up- or down-regulated in colon normal mucosa

p-value corrected for multiple testing as reported in: Lee HK et al. "ErmineJ: Tool for functional analysis of gene expression data sets" BMC Bioinformatics 2005; 6:269

Table 2. Gene with markedly altered expression between ileum and colon normal mucosa (fold-differences of ≥ 4.0) (Statistical significance was assessed with the Mann-Whitney test using a False Discovery Rate of 1% as a threshold)

GENES MORE HIGHLY EXPRESSED IN COLON NORMAL MUCOSA

Affymetrix ID	Fold Change	Gene Symbol	GenBank	Gene Name	Gene Ontology (*Biological Processes*)
211040_s_at	396.2	CEACAM7	AF000623	carcinoembryonic antigen-related cell adhesion molecule 7	unknown
209301_at	127.8	CA2	M36532	carbonic anhydrase II	one-carbon compound metabolism
206784_at	127.2	AQP8	NM_001169	aquaporin 8	water transport
220834_at	104.1	MS4A12	NM_017718	membrane-spanning 4-domains, subfamily A, member 12	unknown
220645_at	60.8	FAM55D	NM_017679	family with sequence similarity 55, member D	unknown
205950_s_at	54.6	CA1	NM_001738	carbonic anhydrase I	one-carbon compound metabolism
220724_at	36.1	FLJ21511	NM_025087	hypothetical protein FLJ21511	unknown
213432_at	33.4	MUC5B	AI687108	mucin 5B, oligomeric mucus/gel-forming	unknown
209772_s_at	18.6	CD24	X69397	CD24 molecule	humoral immune response
219669_at	18.5	CD177	NM_020406	CD177 molecule	unknown
206208_at	18.4	CA4	NM_000717	carbonic anhydrase IV	one-carbon compound metabolism
209692_at	15.9	EYA2	U71207	eyes absent homolog 2 (Drosophila)	eye morphogenesis // mesoderm cell fate specification
205403_at	15.6	IL1R2	NM_004633	interleukin 1 receptor, type II	immune response
201491_s_at	11.3	PYGB	NM_002862	phosphorylase, glycogen, brain	carbohydrate metabolism // glycogen catabolism
217109_at	10.9	MUC4	AJ245547	mucin 4, cell surface associated	unknown
206268_at	10.8	LEFTY1	NM_020997	left-right determination factor 1	transforming growth factor beta receptor signaling pathway // cell growth
200884_at	9.3	CKB	NM_001823	creatine kinase, brain	unknown
219491_at	8.9	LRFN4	NM_024036	leucine rich repeat and fibronectin type III domain containing 4	unknown
206418_at	8.7	NOX1	NM_007052	NADPH oxidase 1	electron transport // NADPH metabolism // FADH2 metabolism // proton transport
205440_s_at	8.4	NPY1R	NM_000909	neuropeptide Y receptor Y1	G-protein signaling, coupled to cyclic nucleotide second messenger // G-protein signaling, adenylate
219799_s_at	8.3	DHRS9	NM_005771	dehydrogenase/reductase (SDR family) member 9	unknown
204508_s_at	8.0	CA12	BC001012	carbonic anhydrase XII	unknown
220020_at	7.7	CLCA4	NM_012128	chloride channel, calcium activated, family member 4	chloride transport
205464_at	6.9	SCNN1B	NM_000336	sodium channel, nonvoltage-gated 1, beta (Liddle syndrome)	sodium ion transport // excretion
213693_s_at	6.8	MUC1	AI810869	mucin 1, cell surface associated	unknown
219312_s_at	6.3	ZBTB10	NM_023929	zinc finger and BTB domain containing 10	unknown
222288_at	6.2	PPP4R2	AI004009	Protein phosphatase 4, regulatory subunit 2	unknown
205097_at	6.2	SLC26A2	AI025519	solute carrier family 26 (sulfate transporter), member 2	sulfate transport
219955_at	6.1	L1TD1	NM_019079	LINE-1 type transposase domain containing 1	unknown
221796_at	5.6	NTRK2	AA707199	neurotrophic tyrosine kinase, receptor, type 2	transmembrane receptor protein tyrosine kinase
205154_at	5.6	LRRN5	NM_006338	leucine rich repeat neuronal 5	cell adhesion // signal transduction
205891_at	5.5	ADORA2B	NM_000676	adenosine A2b receptor	activation of MAPK // cellular defense response // G-protein coupled receptor protein signaling pathway //
209576_s_at	5.2	CEACAM5	NM_001712	carcinoembryonic antigen-related cell adhesion molecule 1 (biliary glycoprotein)	immune response
211401_s_at	5.1	FGFR2	AB030078	fibroblast growth factor receptor 2	unknown
205980_s_at	5.1	ARHGAP8	NM_015388	Rho GTPase activating protein 8	unknown
207432_at	5.1	VMD2L1	NM_017682	vitelliform macular dystrophy 2-like 1	unknown
202236_s_at	5.0	SLC16A1	NM_003051	solute carrier family 16 (monocarboxylic acid transporters), member 1	mevalonate transport
204607_at	4.9	HMGCS2	NM_005518	3-hydroxy-3-methylglutaryl-Coenzyme A synthase 2 (mitochondrial)	acetyl-CoA metabolism // cholesterol biosynthesis
213435_at	4.9	SATB2	AB020957	SATB family member 2	unknown
219103_at	4.8	DDEFL1	NM_017707	development and differentiation enhancing factor-like 1	metabolism
210717_s_at	4.7	LEPREL1	NM_018192	leprecan-like 1	protein metabolism
207653_at	4.6	FOXD2	NM_004474	forkhead box D2	regulation of transcription, DNA-dependent
214433_s_at	4.6	SELENBP1	NM_003944	selenium binding protein 1	unknown
219888_s_at	4.5	NETO2	NM_018092	neuropilin (NRP) and tolloid (TLL)-like 2	unknown
210809_s_at	4.5	POSTN	D13665	periostin, osteoblast specific factor	skeletal development // cell adhesion
215094_at	4.5	PTGDR	AI460323	prostaglandin D2 receptor (DP)	G-protein coupled receptor protein signaling pathway
208510_s_at	4.4	PPARG	NM_015869	peroxisome proliferative activated receptor, gamma	energy pathways // regulation of transcription from Pol II promoter // lipid metabolism // signal
202587_s_at	4.2	AK1	BC001116	adenylate kinase 1	ATP metabolism
220786_s_at	4.2	SLC38A4	NM_018018	solute carrier family 38, member 4	unknown
205185_at	4.2	SPINK5	NM_006846	serine peptidase inhibitor, Kazal type 5	antimicrobial humoral response // anti-inflammatory
214463_x_at	4.1	HIST1H4K	NM_003541	histone 1, H4k	unknown
202742_s_at	4.0	PRKACB	NM_002731	protein kinase, cAMP-dependent, catalytic, beta	unknown

GENES MORE HIGHLY EXPRESSED IN ILEUM NORMAL MUCOSA

Affymetrix ID	Fold Change	Gene Symbol	GenBank	Gene Name	Gene Ontology (*Biological Processes*)
210445_at	2150.5	FABP6	U19869	fatty acid binding protein 6, ileal (gastrotropin)	lipid metabolism // transport // negative regulation of
217073_x_at	645.2	APOA1	X02162	apolipoprotein A-I	lipid transport // circulation // cholesterol metabolism
205820_s_at	606.1	APOC3	NM_000040	apolipoprotein C-III	lipid transport // lipid catabolism
204701_s_at	370.4	ALDOB	BF195990	aldolase B, fructose-bisphosphate	fructose metabolism // glycolysis
206522_at	297.6	MGAM	NM_004668	maltase-glucoamylase (alpha-glucosidase)	carbohydrate metabolism // starch catabolism
211427_s_at	257.7	KCNJ13	AB013889	potassium inwardly-rectifying channel, subfamily J, member 13	unknown
206804_at	242.7	APOA4	NM_000482	apolipoprotein A-IV	lipid metabolism // lipid transport // circulation
208535_at	214.6	SLC2A2	NM_000340	solute carrier family 2 (facilitated glucose transporter), member	carbohydrate metabolism // carbohydrate transport //
205960_at	201.2	AADAC	NM_001086	arylacetamide deacetylase (esterase)	metabolism
208367_x_at	193.4	CYP3A4	NM_000776	cytochrome P450, family 3, subfamily A, polypeptide 4	lipid metabolism // oncogenesis // p450; electron
210165_at	186.2	DNASE1	M55983	deoxyribonuclease I	DNA catabolism // apoptosis
207140_at	142.5	ALPI	NM_001631	alkaline phosphatase, intestinal	alk_phosphatase metabolism
205108_s_at	128.4	APOB	NM_000384	apolipoprotein B (including Ag(x) antigen)	lipid transport // signal transduction // circulation //
204845_s_at	122.2	ENPEP	NM_001977	glutamyl aminopeptidase (aminopeptidase A)	proteolysis and peptidolysis // cell-cell signaling // cell
203435_s_at	99.0	MME	NM_007287	membrane metallo-endopeptidase (neutral endopeptidase, enkephalinase, CALLA, CD10)	proteolysis and peptidolysis // cell-cell signaling
205675_at	80.6	FST	AI823301	FST	lipid metabolism // lipid transport
209952_at	74.1	G6PC	NM_000151	glucose-6 phosphatase, catalytic (glycogen storage disease	unknown
217564_s_at	67.1	CPS1	W80357	carbamoyl-phosphate synthetase 1, mitochondrial	urea cycle // arginine biosynthesis // pyrimidine base
220135_s_at	59.9	SLC7A9	NM_014270	solute carrier family 7 (cationic amino acid transporter, y+	protein complex assembly // amino acid metabolism
204997_at	50.5	GPD1	NM_005276	glycerol-3-phosphate dehydrogenase 1 (soluble)	carbohydrate metabolism // glycerol-3-phosphate
206291_at	50.3	NTS	NM_006183	neurotensin	signal transduction
209937_at	47.6	TM4SF4	BC001386	transmembrane 4 L six family member 4	N-linked glycosylation // negative regulation of cell
207014_at	42.9	DEFA6	NM_001926	defensin, alpha 6, Paneth cell-specific	xenobiotic metabolism // defense response //
205983_at	41.2	DPEP1	NM_004413	dipeptidase 1 (renal)	unknown
207095_at	39.7	SLC10A2	NM_000452	solute carrier family 10 (sodium/bile acid cotransporter family),	sodium ion transport // organic anion transport
205866_at	38.8	FMO1	NM_002021	flavin containing monooxygenase 1	electron transport
203924_at	36.4	GSTA1	NM_000846	glutathione S-transferase A1	unknown
205844_at	36.0	VNN1	NM_004666	vanin 1	nitrogen metabolism // cell motility
205815_at	33.2	REG3A	NM_002580	regenerating islet-derived 3 alpha	acute-phase response // inflammatory response // heterophilic cell adhesion // development // cell
204041_at	32.9	MAOB	NM_000898	monoamine oxidase B	electron transport
219614_s_at	30.1	SLC6A20	NM_020208	solute carrier family 6 (proline IMINO transporter), member 20	neurotransmitter transport
220503_at	29.8	SLC13A1	AF260824	solute carrier family 13 (sodium/sulfate symporters), member 1	sodium ion transport // sulfate transport
207805_at	27.7	NAALAD1.1	NM_005468	N-acetylated alpha-linked acidic dipeptidase-like 1	proteolysis and peptidolysis
217809_s_at	25.5	CYBRD1	NM_024843	cytochrome b reductase 1	electron transport
203716_s_at	25.0	DPP4	M80536	dipeptidyl-peptidase 4 (CD26, adenosine deaminase	proteolysis and peptidolysis // immune response
206988_at	24.6	CCL25	NM_005624	chemokine (C-C motif) ligand 25	chemotaxis // inflammatory response // immune response // G-protein coupled receptor protein
211028_s_at	23.3	KHK	BC008233	ketohexokinase (fructokinase)	unknown

209347_s_at	22.9	**MAF**	BF508646	v-maf musculoaponeurotic fibrosarcoma oncogene homolog (avian)	regulation of transcription, DNA-dependent // transcription from Pol II promoter // cell growth
210324_at	22.0	**C8G**	M17263	complement component 8, gamma polypeptide	transport // complement activation // cytolysis
207251_at	21.4	**SLC15A1**	NM_005073	solute carrier family 15 (oligopeptide transporter), member 1	small molecule transport // digestion
216910_at	21.2	**XPNPEP2**	AF195953	X-prolyl aminopeptidase (aminopeptidase P) 2, membrane-	unknown
205886_at	20.3	**REG1B**	NM_006507	regenerating islet-derived 1 beta (pancreatic stone protein,	heterophilic cell adhesion // cell proliferation
222102_at	20.3	**GSTA3**	NM_000847	glutathione S-transferase A3	response to stress // metabolism
210452_x_at	19.8	**CYP4F2**	D26480	cytochrome P450, family 4, subfamily F, polypeptide 2	p450; electron transport // leukotriene metabolism
209752_at	18.8	**REG1A**	AF172331	regenerating islet-derived 1 alpha (pancreatic stone protein,	unknown
220756_s_at	18.8	**GPR172B**	NM_017986	G protein-coupled receptor 172B	unknown
205670_at	18.0	**GAL3ST1**	NM_004861	galactose-3-O-sulfotransferase 1	N-linked glycosylation
204515_at	18.1	**HSD3B1**	NM_000862	hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid	C21-steroid hormone biosynthesis
209155_at	17.9	**ABCC2**	NM_000392	ATP-binding cassette, sub-family C (CFTR/MRP), member 2	transport
208864_at	17.9	**SI**	NM_001041	sucrase-isomaltase (alpha-glucosidase)	carbohydrate metabolism
206514_s_at	17.6	**CYP4F3**	D12620	cytochrome P450, family 4, subfamily F, polypeptide 3	electron transport // leukotriene metabolism
207249_s_at	17.4	**SLC28A2**	NM_004212	solute carrier family 28 (sodium-coupled nucleoside transporter), member 2	nucleobase, nucleoside, nucleotide and nucleic acid metabolism // purine nucleoside transport
211756_s_at	17.2	**PNLIPRP2**	BC005909	pancreatic lipase-related protein 2 /// pancreatic lipase-related	triacylglycerol metabolism // lipid catabolism
220639_at	16.8	**TM4SF20**	NM_024795	transmembrane 4 L six family member 20	unknown
205380_at	16.4	**PDZK1**	NM_002614	PDZ domain containing 1	unknown
220611_at	16.2	**DAB1**	NM_021080	disabled homolog 1 (Drosophila)	development // neurogenesis
214200_s_at	16.1	**PRODH**	AA074145	proline dehydrogenase (oxidase) 1	proline metabolism
206626_at	16.0	**SLC5A1**	NM_000343	solute carrier family 5 (sodium/glucose cotransporter), member	sodium ion transport // sugar utilization
207378_at	15.9	**TREH**	NM_007180	trehalase (brush-border membrane glycoprotein)	carbohydrate metabolism // trehalose catabolism
207251_at	14.8	**MEP1B**	NM_005925	meprin A, beta	proteolysis and peptidolysis // digestion
208748_s_at	13.8	**ACE**	AI823989	angiotensin I converting enzyme (peptidyl-dipeptidase A) 1	circulation // regulation of blood pressure
214380_at	13.7	**DKFZp564G223**	AI733515	MRNA; cDNA DKFZp564G223 (from clone DKFZp564G223)	unknown
205402_x_at	13.6	**PRSS2**	NM_002770	protease, serine, 2 (trypsin 2)	proteolysis and peptidolysis // digestion
209920_at	13.5	**SLC6A6**	NM_003043	solute carrier family 6 (neurotransmitter transporter, taurine),	amino acid metabolism // neurotransmitter transport
209383_at	13.4	**ABCG5**	NM_022436	ATP-binding cassette, sub-family G (WHITE), member 5	transport // cholesterol absorption
206775_at	13.3	**CUBN**	NM_001081	cubilin (intrinsic factor-cobalamin receptor)	physiological process // vitamin B12 transport
219982_at	12.9	**ACE2**	NM_021804	angiotensin I converting enzyme (peptidyl-dipeptidase A) 2	unknown
207529_at	12.4	**DEFA5**	NM_021010	defensin, alpha 5, Paneth cell-specific	xenobiotic metabolism // defense response //
204430_s_at	12.0	**SLC2A5**	NM_003039	solute carrier family 2 (facilitated glucose/fructose transporter),	unknown
206293_at	11.9	**SULT2A1**	U08024	sulfotransferase family, cytosolic, 2A, dehydroepiandrosterone	digestion // steroid metabolism // bile acid catabolism
207468_s_at	11.8	**SFRP5**	NM_003015	secreted frizzled-related protein 5	apoptosis // establishment and/or maintenance of cell polarity // signal transduction // vision //
220974_x_at	11.3	**SFXN3**	NM_030971	sideroflexin 3	cation transport
205429_s_at	10.9	**MPP6**	NM_016447	membrane protein, palmitoylated 6 (MAGUK p55 subfamily	protein complex assembly
219271_at	10.9	**GALNT14**	NM_024572	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-	unknown
39248_at	10.9	**AQP3**	N74607	aquaporin 3 (Gill blood group)	transport // excretion
211843_x_at	10.8	**CYP3A7**	AF315325	cytochrome P450, family 3, subfamily A, polypeptide 7	electron transport
203946_s_at	10.6	**ARG2**	U75667	arginase, type II	urea cycle // arginine catabolism // nitric oxide
204591_x_at	10.5	**APOC2**	NM_000483	apolipoprotein C-II	lipid transport // circulation // lipid catabolism
210123_s_at	10.2	**CHRNA7**	U62436	cholinergic receptor, nicotinic, alpha 7	unknown
210289_at	9.8	**NAT8**	AB013094	N-acetyltransferase 8 (camello like)	response to drug
210451_at	9.4	**PKLR**	M15465	pyruvate kinase, liver and RBC	unknown
207498_s_at	9.3	**CYP2D6**	NM_000106	cytochrome P450, family 2, subfamily D, polypeptide 6	p450; electron transport
203605_s_at	9.2	**DFNA5**	NM_004403	deafness, autosomal dominant 5	hearing
204855_at	9.2	**SERPINB5**	NM_002639	serpin peptidase inhibitor, clade B (ovalbumin), member 5	unknown
207515_at	9.1	**SLC6A4**	NM_001045	solute carrier family 6 (neurotransmitter transporter, serotonin),	neurotransmitter uptake // serotonin transport
207838_at	8.7	**PRSS7**	NM_002772	protease, serine, 7 (enterokinase)	proteolysis and peptidolysis
206964_at	8.6	**CML2**	NM_016347	putative N-acetyltransferase Camello 2	unknown
211003_x_at	8.5	**TGM2**	BC003551	transglutaminase 2 (C polypeptide, protein-glutamine-gamma-	peptide cross-linking
213230_at	8.3	**CDR2L**	AI422305	cerebellar degeneration-related protein 2-like	regulation of transcription, DNA-dependent
219398_at	8.3	**CIDEC**	NM_022094	cell death-inducing DFFA-like effector c	apoptosis // DNA damage response, signal transduction resulting in induction of apoptosis
204566_s_at	8.1	**SLC7A7**	NM_003982	solute carrier family 7 (cationic amino acid transporter, y+	protein complex assembly // amino acid metabolism
207200_at	8.1	**OTC**	NM_000531	ornithine carbamoyltransferase	urea cycle // arginine biosynthesis
205234_at	7.9	**SLC16A4**	NM_004696	solute carrier family 16 (monocarboxylic acid transporters),	monocarboxylic acid transport
213568_at	7.9	**OSR2**	AI811298	odd-skipped related 2 (Drosophila)	electron transport
210139_s_at	7.8	**PMP22**	L03203	peripheral myelin protein 22	synaptic transmission // peripheral nervous system development // reparative reaulation of cell
221188_s_at	7.6	**CIDEB**	NM_014430	cell death-inducing DFFA-like effector b	apoptosis // DNA damage response, signal transduction resulting in induction of apoptosis
216470_x_at	7.3	**PRSS1**	AF009664	protease, serine, 1 (trypsin 1)	unknown
220070_at	7.2	**JMJD5**	NM_024773	jumonji domain containing 5	unknown
209696_at	7.2	**FBP1**	D26054	fructose-1,6-bisphosphatase 1	carbohydrate metabolism // fructose metabolism //
201599_at	7.1	**OAT**	NM_000274	ornithine aminotransferase (gyrate atrophy)	amino acid metabolism // ornithine metabolism //
210677_at	7.0	**SOAT2**	AF059203	sterol O-acyltransferase 2	cholesterol metabolism
219954_s_at	6.9	**GBA3**	NM_020973	glucosidase, beta, acid 3 (cytosolic)	unknown
210600_at	6.9	**DAK**	NM_015533	dihydroxyacetone kinase 2 homolog (S. cerevisiae)	glycerol metabolism
206422_at	6.9	**GCG**	NM_002054	glucagon	signal transduction // G-protein coupled receptor protein signaling pathway // feeding behavior // cell
220017_x_at	6.6	**CYP2C9**	NM_000771	cytochrome P450, family 2, subfamily C, polypeptide 9	p450; electron transport
202833_s_at	6.5	**SERPINA1**	NM_000295	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase,	acute-phase response
207150_at	6.3	**APOBEC1**	NM_001644	apolipoprotein B mRNA editing enzyme, catalytic polypeptide 1	mRNA editing // mRNA processing // lipid
219640_at	6.3	**CLDN15**	NM_014343	claudin 15	unknown
219956_at	6.2	**C1orf116**	NM_023930	chromosome 1 open reading frame 116	unknown
209213_at	6.2	**CBR1**	BC002511	carbonyl reductase 1	metabolism
213394_at	6.1	**MAPKBP1**	AI674759	mitogen-activated protein kinase binding protein 1	phosphoenolpyruvate-dependent sugar
208450_at	6.1	**LGALS2**	NM_006498	lectin, galactoside-binding, soluble, 2 (galectin 2)	heterophilic cell adhesion
207142_at	6.0	**KCNJ3**	NM_002239	potassium inwardly-rectifying channel, subfamily J, member 3	ion transport // potassium ion transport
216025_x_at	6.0	**CYP2C19**	M21940	cytochrome P450, family 2, subfamily C, polypeptide 19	p450; electron transport
219934_s_at	6.0	**SULT1E1**	NM_005420	sulfotransferase family 1E, estrogen-preferring, member 1	steroid metabolism
213305_at	5.9	**CHN2**	AK026415	chimerin (chimaerin) 2	intracellular signaling cascade
203571_s_at	5.9	**C10orf116**	NM_006829	chromosome 10 open reading frame 116	unknown
209805_at	5.8	**RHOD**	BC001338	ras homolog gene family, member D	Rho protein signal transduction
203669_s_at	5.7	**DGAT1**	NM_012079	diacylglycerol O-acyltransferase homolog 1 (mouse)	triacylglycerol metabolism // fat body development
208186_s_at	5.7	**LIPE**	NM_005357	lipase, hormone-sensitive	energy pathways // protein amino acid phosphorylation // fatty acid metabolism // cholesterol lipid catabolism
203649_s_at	5.5	**PLA2G2A**	NM_000300	phospholipase A2, group IIA (platelets, synovial fluid)	unknown
207901_s_at	5.4	**ESRRG**	NM_001438	estrogen-related receptor gamma	regulation of transcription, DNA-dependent
205081_at	5.4	**CRIP1**	NM_001311	cysteine-rich protein 1 (intestinal)	cell proliferation // antimicrobial humoral response
202100_at	5.4	**PEPD**	NM_000285	peptidase D	amino acid metabolism // collagen catabolism
206768_at	5.3	**EDN2**	NM_001956	endothelin 2	transmembrane receptor protein tyrosine kinase signaling pathway // protein kinase C activation // cell-
208250_s_at	5.3	**DMD11**	NM_004406	deleted in malignant brain tumors 1	unknown
209441_at	5.2	**RHOBTB2**	AY008093	Rho-related BTB domain containing 2	unknown
220078_at	5.0	**EST**	BF110802	Transcribed locus, strongly similar to NP_065940.1	potassium ion transport // sodium ion transport
219334_s_at	5.0	**OBFC2A**	NM_022837	oligonucleotide/oligosaccharide-binding fold containing 2A	unknown
205799_s_at	4.9	**SLC3A1**	M95549	solute carrier family 3 (cystine, dibasic and neutral amino acid transporters, activator of cystine, dibasic and neutral amino	carbohydrate metabolism // amino acid metabolism // basic amino acid transport // L-cystine transport
208741_at	4.9	**C3orf32**	NM_015931	chromosome 3 open reading frame 32	electron transport
221345_at	4.9	**FFAR2**	NM_005306	free fatty acid receptor 2	G-protein coupled receptor protein signaling pathway
217288_s_at	4.9	**SLC37A4**	AF097831	solute carrier family 37 (glycerol-6-phosphate transporter),	unknown
215103_at	4.8	**CYP2C18**	AW162911	cytochrome P450, family 2, subfamily C, polypeptide 18	p450; electron transport
209994_s_at	4.8	**ABCB4**	AF016535	ATP-binding cassette, sub-family B (MDR/TAP), member 4	transport // response to drug
206242_at	4.8	**TM4SF5**	NM_003963	transmembrane 4 L six family member 5	N-linked glycosylation
209992_at	4.7	**ABCB1**	AF016535	ATP-binding cassette, sub-family B (MDR/TAP), member 1	transport // response to drug
218450_at	4.7	**HERP1**	NM_015887	heme binding protein 1	circadian rhythm
201809_at	4.6	**FAM3C**	NM_014060	family with sequence similarity 3, member C	unknown
206340_at	4.6	**NR1H4**	NM_005123	nuclear receptor subfamily 1, group H, member 4	unknown

220186_s_at	4.6	**PCLKC**	NM_017675	protocadherin LKC	homophilic cell adhesion
205825_at	4.6	**PCSK1**	NM_000439	proprotein convertase subtilisin/kexin type 1	proteolysis and peptidolysis // cell-cell signaling //
211442_x_at	4.6	**CYP3A43**	AF280111	cytochrome P450, family 3, subfamily A, polypeptide 43	p450, electron transport
216733_s_at	4.6	**GATM**	X86401	glycine amidinotransferase (L-arginine:glycine	creatine biosynthesis
205715_at	4.6	**BST1**	NM_004334	bone marrow stromal cell antigen 1	humoral immune response // development
202022_at	4.5	**ALDOC**	NM_005165	aldolase C, fructose-bisphosphate	fructose metabolism // glycolysis
212906_at	4.5	**SF3A3**	BE044440	Splicing factor 3a, subunit 3, 60kDa	unknown
206000_at	4.4	**MEP1A**	NM_005588	meprin A, alpha (PABA peptide hydrolase)	proteolysis and peptidolysis // digestion
219786_s_at	4.4	**MUCDHL**	NM_021924	mucin and cadherin-like	unknown
209735_at	4.4	**ABCG2**	AF098951	ATP-binding cassette, sub-family G (WHITE), member 2	transport // response to drug
216316_x_at	4.4	**GK**	X78713	glycerol kinase	unknown
219857_at	4.3	**C10orf81**	NM_024889	chromosome 10 open reading frame 81	unknown
202651_at	4.3	**LPGAT1**	NM_014873	lysophosphatidylglycerol acyltransferase 1	metabolism // phospholipid biosynthesis
205614_x_at	4.3	**MST1**	NM_020998	macrophage stimulating 1 (hepatocyte growth factor-like)	proteolysis and peptidolysis // blood coagulation
206030_at	4.2	**ASPA**	NM_000049	aspartoacylase (Canavan disease)	aspartate catabolism
206396_at	4.2	**SLC4A4**	NM_004170	solute carrier family 1 (neuronal/epithelial high affinity glutamate transporter, system Xag), member 1	dicarboxylic acid transport // synaptic transmission // L-glutamate transport
205898_at	4.2	**SLC22A4**	NM_003059	solute carrier family 22 (organic cation transporter), member 4	unknown
207567_at	4.2	**SLC13A2**	NM_003904	solute carrier family 13 (sodium-dependent dicarboxylate	unknown
219847_at	4.1	**HDAC11**	NM_024827	histone deacetylase 11	regulation of transcription, DNA-dependent //
202357_s_at	4.1	**CFB**	NM_001710	complement factor B	proteolysis and peptidolysis // complement
220292_at	4.1	**ZNF434**	NM_024340	zinc finger protein 434	unknown

Table 3. Genes most likely to be relevant to explain the different incidence of carcinogenesis in ileum and colon (Fold Change ≥ 2; selection from the gene list of Table 2)

Gene Symbol	Gene Name	Fold Differences *	
		Up	Down
Cell Proliferation / Differentiation			
DHRS9	dehydrogenase/reductase (SDR family) member 9	8.3	
MUC1	mucin 1, cell surface associated	6.8	
NTRK2	neurotrophic tyrosine kinase, receptor, type 2	5.6	
CEACAM1	carcinoembryonic antigen-related cell adhesion molecule 1 (biliary glycoprotein)	5.2	
PPARG	peroxisome proliferative activated receptor, gamma	4.4	
SPINK5	serine peptidase inhibitor, Kazal type 5	4.2	
TP53I11	tumor protein p53 inducible protein 11	3.6	
HSPB1	heat shock 27kDa protein 1	3.5	
MYB	v-myb myeloblastosis viral oncogene homolog (avian)	3.5	
E2F1	E2F transcription factor 1	3.2	
BHLHB3	basic helix-loop-helix domain containing, class B, 3	3.1	
PLCE1	phospholipase C, epsilon 1	3.1	
EMP1	epithelial membrane protein 1	3.1	
SFN	stratifin	3.1	
CCL23	chemokine (C-C motif) ligand 23	2.9	
TSPAN1	tetraspanin 1	2.9	
GIPC1	GIPC PDZ domain containing family, member 1	2.8	
TP73L	tumor protein p73-like	2.7	
CTSL2	cathepsin L2	2.6	
TDGF1	teratocarcinoma-derived growth factor 1	2.6	
ADRA2A	adrenergic, alpha-2A-, receptor	2.6	
GPR125	G protein-coupled receptor 125	2.6	
SLPI	secretory leukocyte peptidase inhibitor	2.4	
GMNN	geminin, DNA replication inhibitor	2.4	
EFNB2	ephrin-B2	2.4	
CD9	CD9 molecule	2.4	
VAV3	vav 3 oncogene	2.3	
EHF	ets homologous factor	2.3	
HDAC9	histone deacetylase 9	2.3	
PLAUR	plasminogen activator, urokinase receptor	2.2	
AREG	amphiregulin (schwannoma-derived growth factor)	2.2	
GNPTAB	N-acetylglucosamine-1-phosphate transferase, alpha and beta subunits	2.1	
FGFBP1	fibroblast growth factor binding protein 1	2.1	
NDRG2	NDRG family member 2	2.1	
GPRC5A	G protein-coupled receptor, family C, group 5, member A	2.1	
C6orf66	chromosome 6 open reading frame 66	2.0	
MAP2K6	mitogen-activated protein kinase kinase 6	2.0	
FABP6	fatty acid binding protein 6, ileal (gastrotropin)		2150.5
APOA1	apolipoprotein A-I		645.2
ENPEP	glutamyl aminopeptidase (aminopeptidase A)		122.2
NTS	neurotensin		50.3
TM4SF4	transmembrane 4 L six family member 4		47.6
REG3A	regenerating islet-derived 3 alpha		33.2
DPP4	dipeptidyl-peptidase 4 (CD26, adenosine deaminase complexing protein 2)		25.0
REG1B	regenerating islet-derived 1 beta (pancreatic stone protein, pancreatic thread protein)		20.3
REG1A	regenerating islet-derived 1 alpha (pancreatic stone protein, pancreatic thread protein)		18.8
ABCC2	ATP-binding cassette, sub-family C (CFTR/MRP), member 2		17.9
PDZK1	PDZ domain containing 1		16.4

ANXA13	annexin A13	3.2
SST	somatostatin	3.2
RARRES3	retinoic acid receptor responder (tazarotene induced) 3	3.0
BTG3	BTG family, member 3	2.9
OKL38	pregnancy-induced growth inhibitor	2.9
NEUROD1	neurogenic differentiation 1	2.9
MDK	midkine (neurite growth-promoting factor 2)	2.8
RGS2	regulator of G-protein signalling 2, 24kDa	2.8
PBXIP1	pre-B-cell leukemia transcription factor interacting protein 1	2.5
DBP	D site of albumin promoter (albumin D-box) binding protein	2.5
VIL1	villin 1	2.5
ACHE	acetylcholinesterase (Yt blood group)	2.4
GAS8	growth arrest-specific 8	2.2
KAZALD1	Kazal-type serine peptidase inhibitor domain 1	2.2
ROBO1	roundabout, axon guidance receptor, homolog 1 (Drosophila)	2.2
AZGP1	alpha-2-glycoprotein 1, zinc	2.1
TNFRSF10B	tumor necrosis factor receptor superfamily, member 10b	2.1
NPY	neuropeptide Y	2.0
GTF2I	general transcription factor II, i	2.0
PCAF	p300/CBP-associated factor	2.0
INSIG1	insulin induced gene 1	2.0
BLNK	B-cell linker	2.0

Regulation of Transcription#

SATB2	SATB family member 2	4.9
FOXD2	forkhead box D2	4.6
PPARG	peroxisome proliferative activated receptor, gamma	4.4
TFCP2L1	transcription factor CP2-like 1	3.8
HOXA11	homeobox A11	3.3
E2F1	E2F transcription factor 1	3.2
BHLHB3	basic helix-loop-helix domain containing, class B, 3	3.1
SLC26A3	Solute carrier family 26, member 3	3.1
LITAF	lipopolysaccharide-induced TNF factor	3.0
TP73L	tumor protein p73-like	2.7
OASL	2'-5'-oligoadenylate synthetase-like	2.5
MAZ	MYC-associated zinc finger protein (purine-binding transcription factor)	2.5
FOXA1	forkhead box A1	2.4
EHF	ets homologous factor	2.3
HDAC9	histone deacetylase 9	2.3
SPIB	Spi-B transcription factor (Spi-1/PU.1 related)	2.2
SALL1	sal-like 1 (Drosophila)	2.2
GNPTAB	N-acetylglucosamine-1-phosphate transferase, alpha and beta subunits	2.1
TRIP6	thyroid hormone receptor interactor 6	2.0
ELF3	E74-like factor 3 (ets domain transcription factor, epithelial-specific)	1.8
ZNF323	zinc finger protein 323	1.7
ID3	inhibitor of DNA binding 3, dominant negative helix-loop-helix protein	1.7
MEF2C	MADS box transcription enhancer factor 2, polypep. C (myocyte enhancer factor 2	1.7
KLF4	Kruppel-like factor 4 (gut)	1.6
TGIF	TGFB-induced factor (TALE family homeobox)	1.6
HOXD4	homeobox D4	1.6
HDAC1	histone deacetylase 1	1.6
HOXA7	homeobox A7	1.6
TRIM33	tripartite motif-containing 33	1.5
HOXA10	homeobox A10	1.5

TBX3	T-box 3 (ulnar mammary syndrome)		2.6
PBXIP1	pre-B-cell leukemia transcription factor interacting protein 1		2.5
DBP	D site of albumin promoter (albumin D-box) binding protein		2.5
NR1I2	nuclear receptor subfamily 1, group I, member 2		2.3
LASS2	LAG1 longevity assurance homolog 2 (S. cerevisiae)		2.1
CIITA	class II, major histocompatibility complex, transactivator		2.0
GTF2I	general transcription factor II, i		2.0
PCAF	p300/CBP-associated factor		2.0
TFDP2	transcription factor Dp-2 (E2F dimerization partner 2)		2.0
HOXC10	homeobox C10		1.9
TULP4	tubby like protein 4		1.9
GATA6	GATA binding protein 6		1.9
TBX10	T-box 10		1.9
DIP2C	DIP2 disco-interacting protein 2 homolog C (Drosophila)		1.8
NR5A2	nuclear receptor subfamily 5, group A, member 2		1.8
NCOA1	nuclear receptor coactivator 1		1.8
NFATC3	nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 3		1.8
TRIM32	tripartite motif-containing 32		1.8
NR3C1	nuclear receptor subfamily 3, group C, member 1 (glucocorticoid receptor)		1.7
ZHX3	zinc fingers and homeoboxes 3		1.7
KLF7	Kruppel-like factor 7 (ubiquitous)		1.6
CREB3L2	cAMP responsive element binding protein 3-like 2		1.6
CHES1	checkpoint suppressor 1		1.5
NFIB	nuclear factor I/B		1.5
RARA	retinoic acid receptor, alpha		1.5
ZHX2	zinc fingers and homeoboxes 2		1.5
Wnt signaling pathway			
CD24	CD24 antigen (small cell lung carcinoma cluster 4 antigen)	18.6	
DHRS9	dehydrogenase/reductase (SDR family) member 9	8.3	
POSTN	periostin, osteoblast specific factor	4.5	
PPARG	peroxisome proliferative activated receptor, gamma	4.4	
EMP1	epithelial membrane protein 1	3.1	
EFNB2	ephrin-B2	2.4	
TLE2	transducin-like enhancer of split 2 (E(sp1) homolog, Drosophila)	2.3	
TLE1	transducin-like enhancer of split 1 (E(sp1) homolog, Drosophila)	2.2	
PLAUR	plasminogen activator, urokinase receptor	2.2	
SEMA3C	sema domain, immunoglobulin domain (Ig), basic domain, secreted, (semaphorin)	2.0	
SFRP5	secreted frizzled-related protein 5		11.8
GCG	glucagon		6.8
PLA2G2A	phospholipase A2, group IIA (platelets, synovial fluid)		5.5
EDN2	endothelin 2		5.3
ABCB1	ATP-binding cassette, sub-family B (MDR/TAP), member 1		4.8
CXXC4	CXXC finger 4		3.3
FZD7	frizzled homolog 7 (Drosophila)		2.2
FGF9	fibroblast growth factor 9 (glia-activating factor)		2.0
FZD7	frizzled homolog 7 (Drosophila)		2.0
SLC9A3R1	solute carrier family 9 (sodium/hydrogen exchanger), isoform 3 regulatory factor 1		2.0

* More expressed (Up) or less expressed (Down) in colon than in ileum normal mucosa samples

We have included those genes with a fold change between 2 and 1.5 because even small changes can relevant in these group

Bold: Wnt target genes

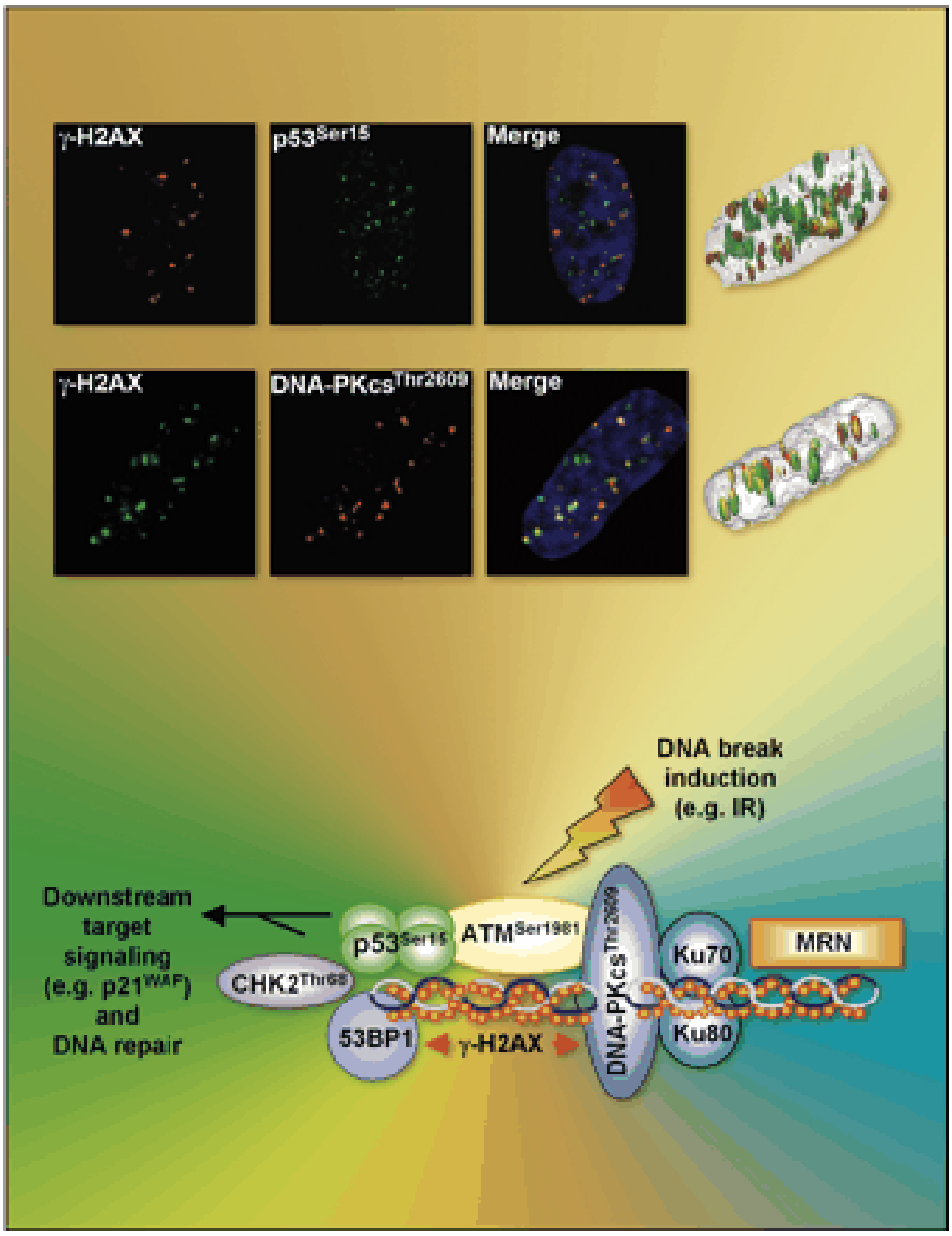
MX1	myxovirus (influenza virus) resistance 1, interferon-inducible protein p78 (mouse)	1.7
TNFRSF1B	tumor necrosis factor receptor superfamily, member 1B	1.6
TNFRSF12A	tumor necrosis factor receptor superfamily, member 12A	1.6
ANXA4	annexin A4	1.6
RTN4	reticulon 4	1.6
TEGT	testis enhanced gene transcript (BAX inhibitor 1)	1.5
BCL2L1	BCL2-like 1	1.5
SNRK	SNF related kinase	1.5
RARA	retinoic acid receptor, alpha	1.5

* More expressed (Up) or less expressed (Down) in colon than in ileum normal mucosa samples

6.5. Appendix V

"Expression of the MutL Homologue hMLH3 in Human Cells and its Role in DNA Mismatch Repair" Elda Cannavo, Giancarlo Marra, Jacob Sabates-Bellver, Mirco Menigatti, Steven M. Lipkin, Franziska Fisher, Petr Cejka and Josef Jiricny. Cancer Research 2005; 65:10759-66.

Cancer Research
A Journal of the American Association for Cancer Research
Volume 65 · Number 23 December 1, 2005 · Pages 10637– 11228



Research Article

Expression of the MutL Homologue hMLH3 in Human Cells and its Role in DNA Mismatch Repair

Elda Cannavo,[1] Giancarlo Marra,[1] Jacob Sabates-Bellver,[1] Mirco Menigatti,[2] Steven M. Lipkin,[3] Franziska Fischer,[1] Petr Cejka,[1] and Josef Jiricny[1]

[1]Institute of Molecular Cancer Research, University of Zurich, Zurich, Switzerland; [2]Department of Internal Medicine, Faculty of Medicine, University of Modena, Modena, Italy; and [3]Departments of Medicine and Biological Chemistry, University of California, Irvine, Irvine, California

Abstract

The human mismatch repair (MMR) proteins hMLH1 and hPMS2 function in MMR as a heterodimer. Cells lacking either protein have a strong mutator phenotype and display microsatellite instability, yet mutations in the *hMLH1* gene account for ~50% of hereditary nonpolyposis colon cancer families, whereas *hPMS2* mutations are substantially less frequent and less penetrant. Similarly, in the mouse model, *Mlh1*$^{-/-}$ animals are highly cancer prone and present with gastrointestinal tumors at an early age, whereas *Pms2*$^{-/-}$ mice succumb to cancer much later in life and do not present with gastrointestinal tumors. This evidence suggested that MLH1 might functionally interact with another MutL homologue, which compensates, at least in part, for a deficiency in PMS2. Sterility of *Mlh1*$^{-/-}$, *Pms2*$^{-/-}$, and *Mlh3*$^{-/-}$ mice implicated the Mlh1/Pms2 and Mlh1/Mlh3 heterodimers in meiotic recombination. We now show that the hMLH1/hMLH3 heterodimer, hMutLγ, can also assist in the repair of base-base mismatches and single extrahelical nucleotides *in vitro*. Analysis of hMLH3 expression in colon cancer cell lines indicated that the protein levels vary substantially and independently of hMLH1. If hMLH3 participates in MMR *in vivo*, its partial redundancy with hPMS2, coupled with the fluctuating expression levels of hMLH3, may help explain the low penetrance of *hPMS2* mutations in hereditary nonpolyposis colon cancer families. (Cancer Res 2005; 65(23): 10759-66)

Introduction

Mismatch repair (MMR) proteins are a highly conserved group of polypeptides that play key roles in the correction of mispairs arising during DNA replication. They also prevent recombination between nonidentical sequences and participate in the signaling of certain types of DNA damage. The importance of MMR proteins in the maintenance of genomic integrity is underscored by the finding that germ line mutations in *MMR* genes predispose to hereditary nonpolyposis colon cancer, a common familial cancer predisposition syndrome (reviewed in refs. 1, 2). The principal MMR players in human cells are homologues of the bacterial MutS and MutL proteins. hMutSα, a heterodimer of the MutS homologues hMSH2 and hMSH6, binds base-base mismatches and small insertion/deletion loops, whereas hMutSβ (a heterodimer of hMSH2 and hMSH3) binds only insertion/deletion loops. This *in vitro* evidence could be corroborated by analysis of the phenotypes of MMR-deficient cells: Those lacking hMSH2 are fully MMR deficient and display a mutator phenotype and microsatellite instability that is consistent with the loss of repair of both base-base mismatches and insertion/deletion loops. Cells lacking hMSH6 retain a strong mutator phenotype but their microsatellite instability is limited to mononucleotide repeats due to the functional redundancy with hMutSβ in insertion/deletion loop repair. This situation is mirrored in hereditary nonpolyposis colon cancer families, where the penetrance of *hMSH2* mutations is substantially higher than that of alterations in the *hMSH6* locus (reviewed in ref. 2).

Whereas it is generally accepted that hMutSα and hMutSβ are the mismatch recognition factors that initiate MMR (reviewed in ref. 3), the function of the MutL homologues remains speculative. The human genome contains numerous genes that have significant sequence homology to *mutL* and to yeast MutL homologue and postmeiotic segregation genes; however, to date, only hMutLα, a heterodimer of hMLH1 and hPMS2, could be shown to be involved in MMR. Correspondingly, hMLH1- or hPMS2-deficient cells have a strong mutator phenotype and high microsatellite instability (reviewed in ref. 1). In *in vitro* studies, hMutLα could be shown to associate with hMutSα on a mismatch-containing substrate (4) and was suggested to act as a "molecular matchmaker" between these protein complexes and the downstream effectors of repair (reviewed in ref. 3). hMutLβ, a heterodimer of hMLH1 and hPMS1, has been biochemically characterized but could not be shown to participate in MMR *in vitro* (5). This finding was substantiated by *in vivo* evidence: Mice carrying a disruption in the *Pms1* gene display neither microsatellite instability nor cancer predisposition (6). hMLH3 was identified through its interaction with hMLH1 on Far Western blots (7); however, this heterodimer, hMutLγ, has not been biochemically characterized and its role in mammalian MMR has not been established. *MLH3* was first identified in *Saccharomyces cerevisiae* and its gene product, scMlh3p, was shown to bind scMlh1p (8, 9) and to be involved in meiotic recombination (reviewed in refs. 10, 11). As *mlh3* mutants display a mutator phenotype similar to that of *msh3*-deficient strains (8, 12), it was suggested that the two polypeptides are involved in the repair of a subset of insertion/deletion loops. hMLH3 seems to be involved in meiotic recombination (13, 14) and the same is true for the murine Mlh3 (14). As both Mlh1- and Mlh3-deficient mice are sterile (reviewed in refs. 10, 11), it was suggested that the two polypeptides function together. However, unlike *Mlh1*$^{-/-}$ animals (6), *Mlh3*$^{-/-}$ mice did not succumb to cancer in the first 9 months of life (15). The roles of the various MMR factors and the phenotypes of mice with defects in *MMR* genes are listed in Table 1.

The involvement of MutL homologue malfunctions in human cancer is not as clear cut as in the case of the MutS homologues. Mutations in *hMLH1* predominate in hereditary nonpolyposis

Requests for reprints: Josef Jiricny, Institute of Molecular Cancer Research, University of Zürich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland. Phone: 411-634-8910; Fax: 411-634-8903; E-mail: jiricny@imcr.unizh.ch.

doi:10.1158/0008-5472.CAN-05-2528

Table 1. Overview of mammalian MutS and MutL homologues and their roles in MMR

Heterodimer	Components	MMR role	Phenotype of knockout mice
hMutSα	hMSH2	Repair of base-base mismatches and small loops	Lymphomas, gastrointestinal, skin, and other tumors
	hMSH6		Lymphomas, gastrointestinal, and other tumors
hMutSβ	hMSH2	Repair of loops	Lymphomas, gastrointestinal, skin, and other tumors
	hMSH3		Gastrointestinal tumors
hMutLα	hMLH1	Repair of all MMR substrates	Lymphomas, gastrointestinal, skin and other tumors; sterility
	hPMS2		Lymphomas, sarcomas; male sterility
hMutLβ	hMLH1	?	Lymphomas, gastrointestinal, skin, and other tumors
	hPMS1		No phenotype
hMutLγ	hMLH1	?	Lymphomas, gastrointestinal, skin, and other tumors
	hMLH3		Sterility

colon cancer, accounting for nearly 50% of all known germ line *MMR* gene mutations (2). Surprisingly, no germ line mutations have been found in *hPMS1* or *hPMS2*, which was unexpected, given the key role of the latter protein in MMR. Recent immunohistochemical analysis of 1,048 unselected colon tumors revealed the lack of hPMS2 in ~1.5%, a proportion similar to that of MSH2-deficient cancers (16). Genetic analysis identified germ line mutations in *hPMS2* in a number of these patients and it is likely that the remainder will also be linked to genetic alterations once the problems associated with sequencing of the *hPMS2* locus are overcome (there are ~20 *hPMS2* pseudogenes on chromosome 7, which interfere with DNA sequencing). However, these patients do not belong to typical hereditary nonpolyposis colon cancer families and the penetrance of these mutations seems to be very low. One possible explanation for this finding is that the defect in hPMS2 is partially compensated for by another MutL homologue, such as hMLH3. Germ line *hMLH3* missense and frameshift mutations have been described in familial colorectal cancer cases but the implication of these alterations in carcinogenesis is ambiguous. In some cases, the mutation in *hMLH3* was identified in families carrying a second *MMR* gene mutation, whereas no mutations in the other *MMR* genes could be identified in other cases (17–19). A similar discrepancy applies also to the microsatellite instability status of the tumors (17, 20). The role of hMLH3 in MMR and of *hMLH3* mutations in cancer thus remains open to question. In an attempt to provide answers to these questions, we examined the role of hMLH3 in MMR *in vitro*.

Materials and Methods

cDNA Vectors

pFastBac1-His$_6$-hMLH3. The cDNA of *hMLH3* (Swiss-Prot entry Q9UHC1) was used as template for a PCR reaction where $(His)_6$ tag was added at the NH_2 terminus of *hMLH3* using the primers hMLH3fo1 (5′-CGCGGATCCACCATGTCGTACTACCATCACCATCACCATCACG-ATTACGATATCCCAACGACCGAAAACCTGTATTTTCAGGGCATCAAGTG-CTTGTCAGTTGAAG-3′) and hMLH3re1 (5′-ATTTGCCTACTGGTGG-GACC-3′). The fragment was then cleaved with *Bam*H1 and *Pfl*M1 and cloned between the corresponding sites in pFastBac1 (Invitrogen, San Diego, CA).

***pTXB1-hMLH3* (amino acids 961-1,453).** The COOH-terminal part of *hMLH3* cDNA coding for amino acids 961 to 1,453 was amplified by PCR from pFastBac1-His$_6$-hMLH3 using the primers fMLH3-Ct (5′-GG-GAATTCCATATGGAGAACTGTGTGATATCAGAAACTC-3′) and rMLH3-Ct (5′-AAGGCCGCTCTTCCGCACATTGGTGGCTCACAGGGAGGCATG-3′). The PCR product was subcloned between the *Nde*I/*Sap*I sites of pTXB1 (New England Biolabs, Beverly, MA).

Expression of hMutLγ

The Bac-to-Bac baculovirus expression system (Life Technologies, Gaithersburg, MD) was used according to the instructions of the manufacturer. *Spodoptera frugiperda* Sf9 cells (2×10^8; Life Technologies) were infected with either a single recombinant baculovirus or with a combination of two viruses at a multiplicity of infection of 10. Cells were harvested 72 hours after infection and total extracts were prepared as described (21). Partial purification of hMutLγ from Sf9 extracts was done using Ni-NTA agarose (Qiagen, Hilden, Germany), and the QIAexpressionist system was used according to the instructions of the manufacturer using 5 mL of 50% Ni-NTA slurry per 100 mg of protein extract.

hMutLγ was expressed also in bacteria using a bicistronic vector *pET11b-His$_6$-hMLH3/MLH1* (cloning information on request) in the BL21 strain of *Escherichia coli*. After induction of expression at 37°C for 4 hours with 0.4 mmol/L isopropyl-β-D-thiogalactopyranoside, the heterodimer was expressed but was insoluble. Nevertheless, the protein could be used to quantify the relative abundance of hMLH3 in HeLa cells.

hMLH3 Antibody Production and Purification

The COOH-terminal polypeptide of hMLH3 (amino acids 961-1,453) was expressed using the Impact-CN-System (New England Biolabs) in BL21 *E. coli* transformed with *pTXB1-hMLH3* (amino acids 961-1,453). The peptide was purified using fast protein liquid chromatography on a MiniQ 4.6/50 PE column (Amersham Pharmacia, Uppsala, Sweden) and used to immunize rabbits at Eurogentec (Seraing, Belgium). The rabbit polyclonal antibody was then affinity-purified using the COOH-terminal polypeptide immobilized on a nitrocellulose membrane. In brief, 100 μg of the purified polypeptide were blotted onto a nitrocellulose membrane by standard electrophoretic transfer, visualized by Ponceau S staining, and the corresponding band was cut out. The membrane was blocked with 5% nonfat dry milk in TBST [20 mmol/L Tris-HCl (pH 7.4), 150 mmol/L NaCl, and 0.1% Tween 20] for 60 minutes, incubated with 700 μL of the polyclonal antibody for 4 hours at 4°C, and washed thrice with TBST for 15 minutes. The membrane was then cut into small pieces (1×0.5 cm) and the antibody was eluted from the membrane by incubation for 20 minutes at room temperature in 0.1 mol/L glycine (pH 2.5). The supernatant was collected and the pH was neutralized by an equal volume of 1 mol/L Tris-HCl (pH 8.0). The purified antibody was stored at −20°C in 50% glycerol.

It was used to perform all the experiments described in this study except for the immunoprecipitation of hMLH3 from human cell extracts.

Human Cell Lines and Preparation of Cell Extracts

All the colon cancer cell lines, HEK293, and HeLa cell lines used in this study were obtained from the cell line repository of Cancer Network Zurich. The hPMS2-deficient cell lines HeLa clone 12 (22) and Hec-1A (23) were kindly provided by Dr. Margherita Bignami (ISS, Rome, Italy). The cell line HEK293T was derived from HEK293 by immortalization with adenovirus 5 DNA and transfection with SV40 large T antigen (24). The *hMLH1* gene in this cell line is epigenetically silenced by promoter hypermethylation (25). The 293T Lα cell line was developed in our laboratory (26). In these cells, the *hMLH1* c-DNA was stably integrated under the control of the tetracycline response promoter

using the Tet-Off system (Clontech, Palo Alto, CA). In the absence of doxycycline, these cells express hMLH1 and are MMR proficient. All the cell lines were cultured at 37°C in a 5% CO_2-humidified atmosphere and maintained in the appropriate media. Whole cell extracts from these cell lines were prepared as described (26) without modifications. The origin and MMR status of the cell lines used in this study is listed in Table 2.

Western Blot Analysis

Western blots were done as previously described (26) using the following primary antibodies: our rabbit polyclonal anti-hMLH3 (1:400), anti-hMLH1 and anti-hPMS2 from BD PharMingen (San Diego, CA) (1:4,000 and 1:1,000, respectively), and anti-β-tubulin (1:2,000; Santa Cruz Biotechnology, Santa Cruz, CA).

Coimmunoprecipitation Analysis of hMLH1 and hMLH3

HeLa whole cell extract (1 mg) was incubated in a total volume of 500 μL in NP40 Lysis Buffer [50 mmol/L Tris-HCl (pH 8.0), 125 mmol/L NaCl, 1% NP40, 2 mmol/L EDTA, 1 mmol/L phenylmethylsulfonyl fluoride, 1× complete protease inhibitory cocktail (Roche Molecular Biochemicals, Basel, Switzerland)] for 3 hours at 4°C with the anti-hMLH1 (6 μg; BD PharMingen) or anti-hMLH3 (10 μg; Santa Cruz Biotechnology) antibodies. The immunoprecipitates were captured by incubation for 30 minutes at 4°C with 50 μL of 50% slurry of Protein A/G PLUS agarose (Santa Cruz Biotechnology). The agarose beads were then washed thrice with cold NP40 Lysis Buffer and the proteins were eluted with SDS sample buffer and subjected to Western blot analysis. Control experiments were done either in the absence of antibody or in the presence of 25 units of Benzonase (Merck, Whitehouse Station, NJ).

Analysis of the *hMLH3* Promoter and Treatment of Cells with 5-Aza-2′-deoxycytidine

The *hMLH3* 5′ flanking region was analyzed for CpG content with the CpG plot software of the European Bioinformatics Institute (http://www.ebi.ac.uk/emboss/cpgplot/) and its methylation status was evaluated with methylation-specific PCR as described previously (16). Primer sequences for the unmethylated reactions were 5′-GTTGTGTGTAGTTTTTGGAGTTG-3′ (sense) and 5′-CTCCCAACACCTAAAACTAACA-3′ (antisense), which amplified a 229 bp product. The methylation-specific primers were 5′-CGCGTAGTTTTCGGAGTC-3′ (sense) and 5′ CTAAAACTAACGAAACGCACG 3′ (antisense), which amplified a 205 bp product. The PCR conditions are available on request.

To reactivate the expression of *hMLH1* and *hMLH3*, 2.5×10^5 HEK293T cells were seeded on a 78 cm^2 dish on day 0 and treated with 3 μg/mL of 5-aza-2′-deoxycytidine (Fluka, Buchs, Switzerland) on days 2 and 5. The medium was changed 24 hours after each addition of the drug and the cells were harvested on day 8.

Microarray Experiments

Microarray experiments were done as described previously (27). Gray columns in the graphs represent mRNA levels based on raw signals detected in the corresponding cell lines with the Affymetrix HG-U133A microarray.

Mismatch Repair Assays

The assays were done as described previously (28, 29).

Results

Expression of hMutLγ in Sf9 cells and production of anti-hMLH3 antibody. To produce the recombinant hMLH3 and hMutLγ factors, *S. frugiperda* Sf9 cells were infected with baculoviruses carrying cDNAs encoding *hMLH1* and/or *hMLH3*. Infection of Sf9 cells with the *hMLH3* virus alone yielded the protein in an amount that was hardly detectable by Western blotting. The amount of expressed protein was significantly increased when the cells were coinfected with both *hMLH1* and *hMLH3* vectors (Fig. 1*A*), suggesting that the presence of hMLH1 is necessary for the stabilization of hMLH3 in Sf9 cells. This is reminiscent of hMSH6 and hPMS2, both of which require their heterodimeric partners (hMSH2 and hMLH1, respectively) for stability. However, the amount of the recombinant heterodimer obtained was too low

Table 2. Characteristics of the human cell lines used in this study

Cell lines	Origin	MMR status*	MMR protein defect†	Genetic complementation
293	Embryonic kidney epithelium	+		
293T	Embryonic kidney epithelium	−	**hMLH1**, hPMS2, **hMLH3**	
293T Lα+	Embryonic kidney epithelium	+	**hMLH3**	*hMLH1* cDNA
CaCo2	Colon carcinoma	+		
CO115	Colon carcinoma	−	**hMLH1**, hPMS2	
Colo741	Colon carcinoma	+		
CX-1	Colon carcinoma	+		
GP5D	Colon carcinoma	−	**hMSH2**, hMSH6, hMSH3, **hMLH3**	
HCT116	Colon carcinoma	−	**hMLH1**, hPMS2, hMSH3	
HCT116+Ch.3	Colon carcinoma	+	hMSH3	Chromosome 3
HT29	Colon carcinoma	+		
Hec1A	Endometrial adenocarcinoma	−	**hPMS2, hMSH6**	
Hec1A+Ch.7	Endometrial adenocarcinoma	−	**hMSH6**	Chromosome 7
HeLa	Cervical carcinoma	+		
HeLa clone 12	Cervical carcinoma	−	**hPMS2**	
LS411	Colon carcinoma	−	**hMLH1**, hPMS2	
SW48	Colon carcinoma	−	**hMLH1**, hPMS2	
SW480	Colon carcinoma	+		
SW837	Colon carcinoma	+		

*+, MMR proficient; −, MMR deficient (these cell lines are unable to repair both base-base mismatches and insertion/deletion loops, with the exception of Hec1A+Ch.7, which is able to repair insertion/deletion loops).

†The primary alteration of MMR protein expression is reported in bold. Lack of hMLH1 or hMSH2 lead to proteolytic degradation of hPMS2, or hMSH6 and hMSH3, respectively. The *hMSH3* gene in HCT116 cells is mutated as a consequence of the MMR defect. The hMLH3 alterations are those described in this study; other alterations have been reported elsewhere (see text).

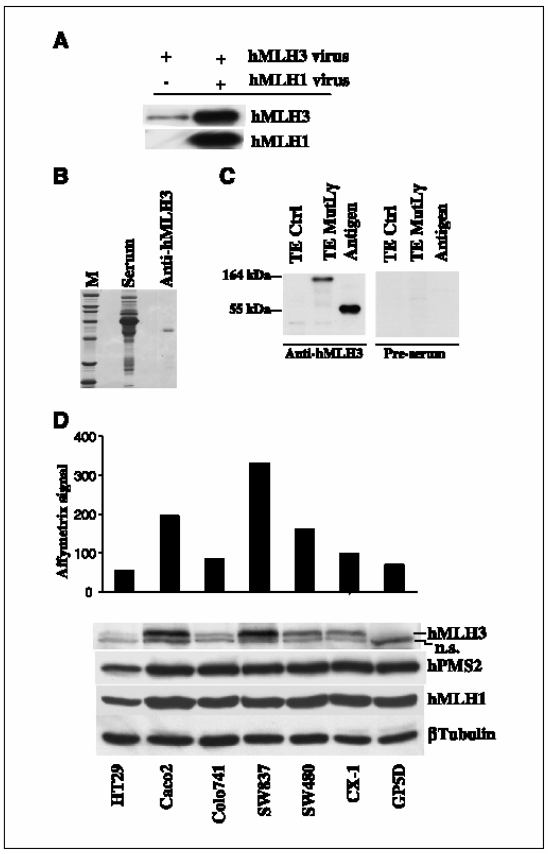


Figure 1. Expression of hMutLγ in Sf9 cells, anti-hMLH3 antibody specificity, and endogenous levels of hMLH3 in colon cancer cell lines. *A*, expression of hMLH3 in Sf9 cells infected either with a baculovirus vector expressing hMLH3 or with a mixture of hMLH1- and hMLH3-expressing viruses. This Western blot shows that hMLH3 is stabilized in this system by hMLH1. *B*, Coomassie blue staining of rabbit polyclonal anti-hMLH3 serum before (*Serum*) and after (*Anti-hMLH3*) affinity purification. *C*, Western blot analysis of whole cell extracts of Sf9 cells uninfected (*TE Ctrl*, 2 μg) or coinfected with the hMLH1/hMLH3 baculoviruses (*TE MutLγ*, 2 μg). The recombinant polypeptide (amino acids 961-1,453) used for the generation of the antibody was loaded in the third lane (1 ng). *Anti-hMLH3*, affinity-purified serum used at 1:400 dilution; *Pre-serum*, preimmune serum used at the same dilution. *D*, microarray analysis of mRNA expression levels (*top*) and Western blot analysis of protein levels (*bottom*) of hMLH3 in a series of colon cancer cell lines (50 μg of whole cell extract per lane); *n.s.*, nonspecific band detected by the anti-hMLH3 antibody in human cell extracts.

to permit extensive purification. The reasons underlying the low levels of expression are unknown at this time, but it is possible that high amounts of the full-length protein may be toxic (7).

Commercially available antibodies could detect the recombinant hMLH3 protein on Western blots but failed to detect the endogenous protein in all the human cell lines used in this study (data not shown). Therefore, we raised our own polyclonal rabbit antiserum, directed against the COOH terminus of hMLH3, which contains the hMLH1-interacting domain. The affinity-purified antibody (Fig. 1*B*) detected a band of the expected size (~160 kDa) in Sf9 lysates infected with the *hMLH1* and *hMLH3* vectors, whereas no signal was visible when we probed lysates of uninfected cells (Fig. 1*C*). The purified antibody was then tested using extracts of various human colon cancer cell lines. The antibody highlighted a double band migrating at the expected size of hMLH3 (Fig. 1*D*, *bottom*). As the faster migrating band was also observed in Western blots done with the preimmune serum (data not shown), and as the abundance of the slower-migrating band correlated with *hMLH3* mRNA expression levels in the same cell lines (Fig. 1*D*, *top*), we concluded that the latter is the specific band. As shown in Fig. 1*D*, the levels of hMLH3 fluctuate significantly in the tested cells lines and seem to be independent of the amount of hMLH1 and hPMS2 expressed in the same cells.

Relative abundance of hMLH3 in human cells and its interaction with hMLH1. Given that hMLH3, hPMS2, and hPMS1 interact with the same region of hMLH1 (30), we wanted to ask whether the relative abundance of the three different heterodimers can be correlated with the phenotype of the cells. Therefore, we did semiquantitative Western blots where we compared the intensity of bands due to endogeneous hMLH3 and hPMS2 proteins in HeLa cells with that of bands due to known amounts of the corresponding recombinant proteins (Fig. 2*A*). These experiments revealed that hMLH3 is ~60 times less abundant than hPMS2. Considering that hPMS1 is ~10 times less abundant than hPMS2 in human cells (5), hMLH3 exists in the cells at levels significantly

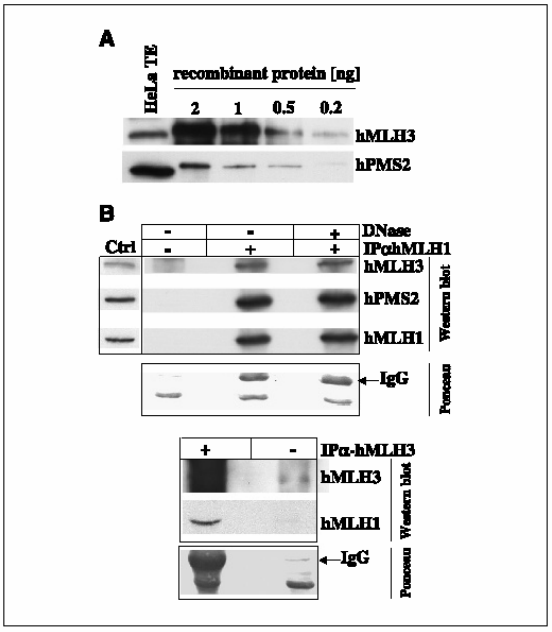


Figure 2. Relative abundance of hMLH3 and its interaction with hMLH1 *in vivo*. *A*, the recombinant hMLH3 and hPMS2 proteins were loaded onto a denaturing polyacrylamide gel in the indicated amounts and visualized by Western blotting with the respective antibodies. The relative abundance of the two polypeptides was calculated by comparing the intensity of the hMLH3 and hPMS2 bands with those of the endogenous proteins present in 50 μg of HeLa whole cell extract. The blot is representative of two independent experiments, and the intensities of the bands were calculated by densitometry. *B*, coimmunoprecipitation of hMLH3 and hMLH1 in HeLa cells. One milligram of whole cell extract was incubated with or without 6 μg of anti-hMLH1 antibody (*top*) or 10 μg of anti-hMLH3 antibody (*bottom*). *DNase*, reaction done in the presence of 25 units of Benzonase. Ponceau staining for IgG is also shown to show equal loading. *Ctrl*, 50 μg of HeLa whole cell extract.

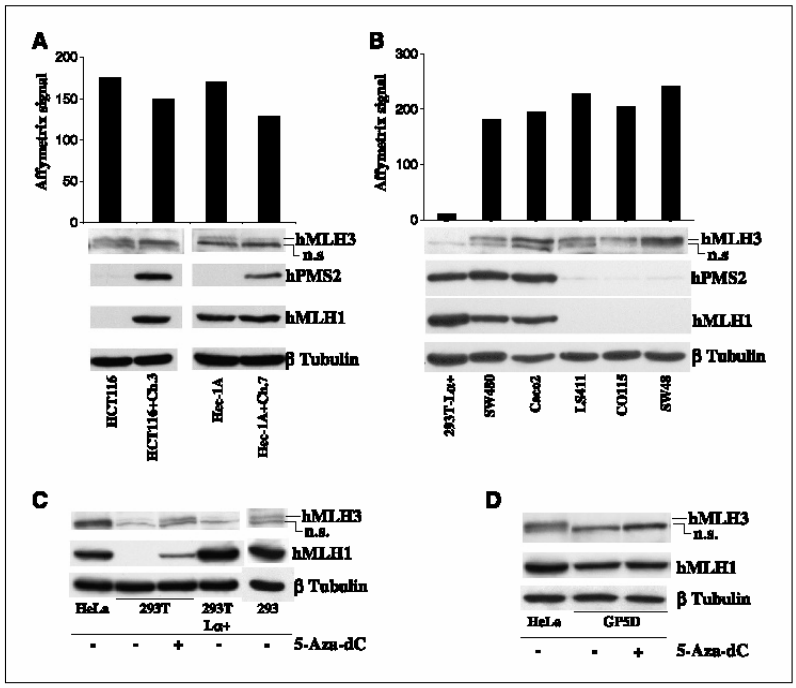


Figure 3. Expression of hMLH3 *in vivo* is independent of hMLH1 and hPMS2 and can be controlled by cytosine methylation. *A*, microarray analysis of *hMLH3* mRNA (*top*) and protein (*bottom*) expression in hMLH1-deficient HCT116 and hPMS2-deficient Hec-1A cells. Correction of the MMR defect by transfer of chromosome 3 or 7, which carry wild-type copies of the *hMLH1* and *hPMS2* genes, respectively, had no effect on hMLH3 expression. *B*, hMLH3 expression is independent of hMLH1/hPMS2 expression in a panel of MMR-proficient and MMR-deficient cell lines. Legend as in (*A*). *C*, the *hMLH3* promoter in 293T cells in silenced by methylation. Expression of *hMLH1* in 293T-Lα cells does not alter hMLH3 levels but demethylation of the promoter by 5-aza-2′-deoxycytidine (*5-Aza-dC*) treatment results in the reappearance of hMLH3, together with hMLH1, the promoter of which is also methylated in these cells. *D*, down-regulation of the *hMLH3* gene in GP5D cells is not mediated by cytosine methylation. The promoter was shown by methylation-specific PCR to be unmethylated and 5-Aza-dC did not reactivate hMLH3 expression in these cells. In the Western blot experiment, 50 μg of total extract were used per lane.

lower than those of the other two hMLH1-interacting partners hPMS2 and hPMS1. In spite of this difference, hMLH3 was found to physically interact with hMLH1 in Far Western experiments (7) and in mammalian two hybrid assays (30). We could confirm this interaction using immunoprecipitation experiments in which the anti-hMLH1 antibody could immunoprecipitate both hMLH3 and hPMS2 from human cell lysates (Fig. 2*B*, *top*) and the anti-hMLH3 antibody precipitated the endogenous hMLH1 (Fig. 2*B*, *bottom*). No proteins were detected in control experiments where the precipitating antibody was omitted. The interaction between hMLH3 and hMLH1 was not mediated by bound DNA because treatment with DNase before incubation with the antibodies failed to abolish the interaction between the two proteins (Fig. 2*B*, *top*).

hMLH1 is not required for the stability of hMLH3 in human cells. hPMS2 and hPMS1 are stabilized by the presence of hMLH1 (1, 31). Considering this characteristic of these two MutL homologues, together with the finding that hMLH1 was required for the stabilization of hMLH3 in baculovirus-infected Sf9 cells (Fig. 1*A*), we expected to observe substantially decreased levels of endogenous hMLH3 in human cell lines lacking hMLH1. Surprisingly, we could detect hMLH3 in hMLH1-deficient HCT116 cells, and the restoration of hMLH1 expression by chromosome 3 transfer resulted in no appreciable increase in hMLH3 level (Fig. 3*A*, *bottom left*). This shows that the presence of hMLH1 is not required for hMLH3 stability in human cells. The relative amounts of intracellular hMLH3 were also unaffected by hPMS2 levels, as shown by comparison of hMLH3 band intensity in Western blots of extracts of the hPMS2-deficient Hec-1A cells with those of a Hec-1A clone in which the expression of hPMS2 was restored by chromosome 7 transfer (Fig. 3*A*, *bottom right*). hMLH3 protein levels failed to correlate with hMLH1 and hPMS2 expression also in other colon cancer cell lines, such as SW480 or Caco2, that express both hMLH1 and hPMS2, or LS411, CO115, or SW48 that lack hMutLα (Fig. 3*B*).

Having established that the level of hMLH3 in cells is not dependent on hMLH1 but that it correlates well with *hMLH3* mRNA levels (Fig. 3*B*), we wondered whether the fluctuation of hMLH3 expression in the tested cell lines could be linked with cytosine methylation, which is known to silence several key genes in colon cancer (32). The human embryonic kidney cell line 293T is deficient in both hMLH1 and hMLH3 (Fig. 3*C*) and it could be shown that the CpG islands that constitute the promoters of hMLH1 (25) and other genes (33) are silenced by hypermethylation in these cells. As the *hMLH3* promoter also contains a CpG island, we reasoned that the lack of *hMLH3* expression in this cell line might also be linked to the transcriptional inactivation of its promoter. This prediction was substantiated in two independent experiments. First, treatment of 293T cells with the demethylating agent 5-aza-2′-deoxycytidine partially restored the expression of both hMLH1 and hMLH3 (Fig. 3*C*). In the second experiment, we treated genomic DNA of 293T and the parental 293 cells (which express both hMLH1 and hMLH3; Fig. 3*C*) with sodium bisulfite, which deaminates cytosines to uracils, but leaves 5-methycytosines unchanged. Methylation-specific PCR showed that the promoter of the *hMLH3* gene in 293T cells was indeed methylated (data not shown). As expected, expression of high amounts of hMLH1 in the 293T-derived 293T Lα+ cells resulted in the stabilization of hPMS2 (26) but did not affect hMLH3 levels (Fig. 3*C*). The promoter of the *hMLH3* gene can thus be silenced by cytosine methylation, but this is most likely not the only mechanism that results in the lack of expression of the protein, as 5-aza-2′-deoxycytidine treatment failed to induce the expression of hMLH3 in GP5D cells (Fig. 3*D*).

Role of hMutLγ in *in vitro* mismatch repair. The observation that extracts from 293T-Lα+ cells are MMR proficient (26) despite their lack of hMLH3 suggested that hMutLγ does not play a major role in MMR *in vitro*. However, the possibility that it acts as a backup to hMutLα in the absence of hPMS2 could not be excluded. Therefore, we tested extracts of the human cell line HeLa clone 12, which expresses hMLH1, hPMS1, and hMLH3 but lacks hPMS2. As shown in Fig. 4*A*, these extracts were deficient in the repair of heteroduplex substrates containing either a G/T mismatch or an insertion/deletion loop of one or two nucleotides, but their repair proficiency on all tested substrates could be restored by the addition of recombinant hMutLα. Before concluding that hMutLγ does not participate in MMR, we considered the possibility that the expression level of endogenous hMLH3 in the tested human cell lines might be too low to be detectably active in our *in vitro* assay. Therefore, we decided to test whether *in vitro* MMR activity may be detected in the presence of higher amounts of the heterodimer. These experiments were done with the hMutL α, β, and γ deficient extracts of 293T cells supplemented with whole cell extracts from Sf9 cells expressing comparable amounts of hMutLα or hMutLγ (Fig. 4*B*, *inset*). As shown previously, extracts of Sf9 cells overexpressing hMutLα could complement the MMR defect in the 293T extracts very efficiently, whereas extracts of uninfected Sf9 cells failed to do so (Fig. 4*B*; ref. 26). Interestingly, when extracts of Sf9 cells expressing hMutLγ were used, we observed an increase in repair activity of ~20%. Similar results were obtained when the hMutLγ was enriched by Ni-agarose chromatography, showing that the observed MMR activity was specific to hMutLγ. We detected similar repair activities on substrates containing a G/T mismatch or a 1-base loop with a nick located either 5′ or 3′ from the mismatch, but no activity was observed on a substrate containing insertion/deletion loops of two or four nucleotides (Fig. 4*B*; data not shown). These experiments show that although physiologic levels of hMutLγ are insufficient to mediate mismatch correction in our *in vitro* MMR assays, the factor can participate, albeit with low efficiency, in the correction of base-base mispairs and one-nucleotide insertion/deletion loops.

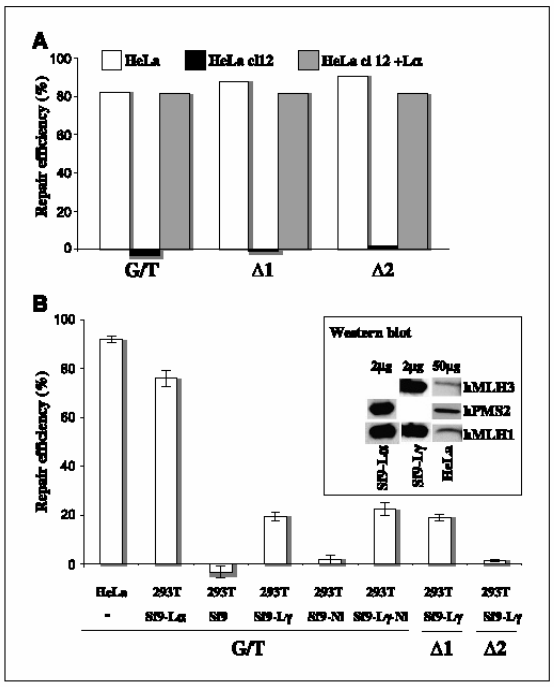


Figure 4. *In vitro* MMR assays. *A*, the MMR defect of cytoplasmic extracts of the hPMS2-deficient cell line HeLa clone 12 can be corrected by the addition of 0.2 μg purified hMutLα. The repair efficiencies were determined on heteroduplex substrates containing a G/T mismatch, or +1 (*Δ1*) or +2 (*Δ2*) insertion/deletion loops. *B*, cytoplasmic extracts of 293T cells, which lack hMLH1, hMLH3, and hPMS2 were supplemented with 2 μg of whole cell extract from Sf9 cells coinfected with hMLH1/hPMS2 (*Sf9-Lα*), hMLH1/hMLH3 (*Sf9-Lγ*), or the latter partially purified on Ni-agarose (*Sf9-Lγ-Ni*). MMR efficiency was tested on heteroduplex substrates containing a G/T mismatch, or +1 (Δ1) or +2 (Δ2) insertion/deletion loops. The amounts of hMutLα and hMutLγ complexes used for complementation were comparable (*inset*). Whole cell extracts from uninfected Sf9 cells (*Sf9*) or Sf9 extract that underwent Ni-agarose (*Sf9-Ni*) chromatography were used as negative controls. *Columns*, result of at least three independent experiments; *bars*, SE. Cytoplasmic extract of the MMR-proficient HeLa cells was used as the positive control.

Discussion

Like its *S. cerevisiae* homologue (8), the mammalian *MLH3* gene (7) could be shown to be involved in meiotic recombination (13–15). However, whereas the *S. cerevisiae* (8) and *Schizosaccharomyces pombe* (12) proteins play a small but distinct role in the repair of a subset of insertion/deletion loops, no similar evidence existed for mammalian MLH3. In this present study, we set out to search for this evidence.

We first wanted to study the expression of hMLH3 and confirm the existence of hMutLγ *in vivo*. Using a newly generated antibody, we showed that hMLH3 is much less abundant than the other two known hMLH1 interactors, hPMS2 and hPMS1. Despite this, we could confirm the physical interaction between hMLH3 and hMLH1 in HeLa cells by immunoprecipitation experiments. Surprisingly, although hMLH1 was required for hMLH3 stability in Sf9 cells (Fig. 1*A*), no such requirement was apparent in human cells where no degradation of hMLH3 occurred in the absence of hMLH1 (Fig. 3*A* and *B*). We also failed to observe any significant competition between hMLH3 and hPMS2 for hMLH1, showing that in human cells hMLH3 might be stabilized by interaction with another, as yet unidentified, protein. This finding is supported by evidence from meiosis in mice, where Mlh3 was seen to bind to pachytene chromosomes before Mlh1 and, after Mlh1 recruitment to these sites, foci containing Mlh3 alone persisted (11, 15). It was, therefore, suggested that Mlh3 could either exist alone or interact with a different partner (11). Immunoprecipitation experiments revealed a direct interaction of *scMlh3p* with *Sgs1* helicase in meiotic *S. cerevisiae* cells (34) and hMLH3 was shown to bind hMSH4 in meiotic human cells (14); however, the identification of the putative hMLH3 partners that might help stabilize it in mitotic cells must await the results of future experiments.

The ultimate objective of this work was to elucidate the role of hMLH3 in human MMR. We first tested extracts of human HeLa clone 12 cells, which lack hPMS2 (22) and thus contain only hMutLβ and hMutLγ. As the former heterodimer is devoid of MMR activity in our *in vitro* MMR assay (31), any observed repair activity could be ascribed to hMutLγ. The extracts were MMR deficient on all tested substrates (Fig. 4*A*), which suggested that the hMutLγ heterodimer does not participate in MMR. However, as hMLH3 is generally much less abundant in human cells than hPMS2, we

wanted to exclude the possibility that the lack of repair activity is linked to insufficient amounts of hMutLγ. Therefore, we tested the MMR activity of extracts of 293T cells, which are deficient in all three MutL homologues, supplemented either with recombinant hMutLα or hMutLγ (Fig. 4*B*). The former factor complemented the MMR defect in the 293T extracts on all tested substrates. When comparable amounts of hMutLγ were used, we observed a small but significant (~20%) repair with both G/T and +1 insertion/deletion loop substrates. This repair activity was not due to an intrinsic repair activity of the Sf9 extracts per se, as extracts from uninfected Sf9 cells were repair deficient in the complementation experiments. As there are no available functional assays to test the activity of hMutLγ, a possibility exists that this heterodimer was isolated in a partially inactive form. However, we consider this possibility unlikely because all the procedures used were identical to those used for the preparation of the Sf9 extract expressing hMutLα, which was fully active. Moreover, immunoprecipitation experiments done with Sf9 extracts expressing hMutLγ showed that hMLH3 was able to bind hMLH1 (data not shown). The sensitivity of the *in vitro* MMR assay remains, however, rather low so that the contribution of hMutLγ to the repair process *in vivo* might be higher. Interestingly, the repair activity of hMutLγ was limited to G/T mismatch and 1-base loops, as we failed to observe any repair activity using +2- and +4-base-loop substrates. The latter result indicates that hMutLγ seems to be involved in the repair of substrates recognized by hMutSα rather than insertion/deletion loops of more than one extrahelical nucleotide recognized by hMutSβ. This is in contrast to the data obtained in *S. cerevisiae* where the role of *scMlh3p* seems to be in the repair of a subset of insertion/deletion loops together with *scMutSβ*. The role of hMLH3 in mammals thus might differ from that in lower eukaryotes.

Our findings, suggesting that hMutLγ may play a backup role in human MMR, are supported by evidence from the mouse model. As noted above, Mlh3 *null* mice were not cancer prone in the first 9 months of life and showed no gross defects in MMR (15). However, a long-term study of these animals, coupled with a highly sensitive analysis of their genomic DNA, provides evidence for the involvement of Mlh3 defects in both MMR and tumorigenesis. $Mlh3^{-/-}$ mice have a shorter life span than the wild-type controls and more than half of the animals develop cancers, including gastrointestinal tumors after the 9-month time span. Importantly, Mlh3 deficiency increased the levels of mutations in long mononucleotide repeats, although to a lesser extent than in $Pms2^{-/-}$ mice (35). Taken together, our results and the mouse model data suggest that the hMutLγ heterodimer functions in the repair of base-base mismatches and small insertion/deletion loops.

Considering the possible involvement of hMLH3 in human MMR, the identification of hMLH3 silencing through promoter hypermethylation is of particular interest. We showed that the hMLH3 promoter is methylated in 293T cells and that the protein is consequently not expressed. In this particular cell line, the methylation could be caused by the presence of the SV40 large T antigen. However, using methylation-specific PCR, we could detect partially methylated hMLH3 promoters in the colon cancer cell line LS411 and in the ovarian cancer cell line A2780/CP70, and fully methylated in the leukemia cell line Jurkat (data not shown), which shows that hMLH3 silencing via promoter hypermethylation can also be unrelated to the presence of SV40 large T antigen.

Although recombinant hMutLγ possessed detectable repair activity in our *in vitro* MMR assays, hPMS2-deficient cells expressing hMLH3 display a strong mutator phenotype (refs. 16, 23; this study). This suggests that hMLH3, most likely in the form of hMutLγ, does not play a major role in MMR *in vivo*. However, the detection of sequence variants of hMLH3 in the germ line of families predisposed to colorectal cancer (17, 20), coupled with our detection of epigenetic silencing of hMLH3 in human cell lines, suggests that this gene may play a role in human cancer, possibly in combination with other risk factors. If hMutLγ does indeed play a backup role for hMutLα *in vivo*, the fluctuating abundance of hMLH3, such as that observed in the tested cell lines (Figs. 1 and 3), might help explain the variable penetrance of *hPMS2* mutations in hereditary nonpolyposis colon cancer families (16).

Acknowledgments

Received 7/21/2005; revised 8/25/2005; accepted 9/19/2005.

Grant support: Swiss Bridge (P. Cejka and J. Jiricny), Swiss National Science Foundation grant 3100/068182.02 (E. Cannavo and J. Jiricny), Union Bank of Switzerland AG (F. Fischer and J. Jiricny), and Swiss Cancer League (J. Sabatés-Bellver and G. Marra).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

We thank Christine Hemmerle for technical assistance and Dr. Pavel Janscak for help with protein purification.

References

1. Jiricny J, Marra G. DNA repair defects in colon cancer. Curr Opin Genet Dev 2003;13:61–9.
2. Peltomaki P, Vasen H. Mutations associated with HNPCC predisposition—update of ICG-HNPCC/INSiGHT mutation database. Dis Markers 2004;20:269–76.
3. Kunkel TA, Erie DA. DNA mismatch repair. Annu Rev Biochem 2005;74:681–710.
4. Raschle M, Dufner P, Marra G, Jiricny J. Mutations within the hMLH1 and hPMS2 subunits of the human MutLα mismatch repair factor affect its ATPase activity, but not its ability to interact with hMutSα. J Biol Chem 2002;277:21810–20.
5. Räschle M, Marra G, Nyström-Lahti M, Schär P, Jiricny J. Identification of hMutLβ, a heterodimer of hMLH1 and hPMS1. J Biol Chem 1999;274:32368–75.
6. Prolla TA, Baker SM, Harris AC, et al. Tumour susceptibility and spontaneous mutation in mice deficient in Mlh1, Pms1 and Pms2 DNA mismatch repair. Nat Genet 1998;18:276–9.
7. Lipkin SM, Wang V, Jacoby R, et al. MLH3: a DNA mismatch repair gene associated with mammalian microsatellite instability. Nat Genet 2000;24:27–35.
8. Flores-Rozas H, Kolodner RD. The *Saccharomyces cerevisiae* MLH3 gene functions in MSH3-dependent suppression of frameshift mutations. Proc Natl Acad Sci U S A 1998;95:12404–9.
9. Wang TF, Kleckner N, Hunter N. Functional specificity of MutL homologs in yeast: evidence for three Mlh1-based heterocomplexes with distinct roles during meiosis in recombination and mismatch correction. Proc Natl Acad Sci U S A 1999;96:13914–9.
10. Hoffmann ER, Borts RH. Meiotic recombination intermediates and mismatch repair proteins. Cytogenet Genome Res 2004;107:232–48.
11. Kolas NK, Cohen PE. Novel and diverse functions of the DNA mismatch repair family in mammalian meiosis and recombination. Cytogenet Genome Res 2004;107:216–31.
12. Harfe BD, Minesinger BK, Jinks-Robertson S. Discrete *in vivo* roles for the MutL homologs Mlh2p and Mlh3p in the removal of frameshift intermediates in budding yeast. Curr Biol 2000;10:145–8.
13. Marcon E, Moens P. MLH1p and MLH3p localize to precociously induced chiasmata of okadaic-acid-treated mouse spermatocytes. Genetics 2003;165:2283–7.
14. Santucci-Darmanin S, Neyton S, Lespinasse F, Saunieres A, Gaudray P, Paquis-Flucklinger V. The DNA mismatch-repair MLH3 protein interacts with MSH4 in meiotic cells, supporting a role for this MutL homolog in mammalian meiotic recombination. Hum Mol Genet 2002;11:1697–706.
15. Lipkin SM, Moens PB, Wang V, et al. Meiotic arrest and aneuploidy in MLH3-deficient mice. Nat Genet 2002;31:385–90.
16. Truninger K, Menigatti M, Luz J, et al. Immunohistochemical analysis reveals high frequency of PMS2 defects in colorectal cancer. Gastroenterology 2005;128:1160–71.

17. Liu HX, Zhou XL, Liu T, et al. The role of hMLH3 in familial colorectal cancer. Cancer Res 2003;63:1894–9.
18. Hienonen T, Laiho P, Salovaara R, et al. Little evidence for involvement of MLH3 in colorectal cancer predisposition. Int J Cancer 2003;106:292–6.
19. de Jong MM, Hofstra RM, Kooi KA, et al. No association between two MLH3 variants (S845G and P844L) and colorectal cancer risk. Cancer Genet Cytogenet 2004;152:70–1.
20. Wu Y, Berends MJ, Sijmons RH, et al. A role for MLH3 in hereditary nonpolyposis colorectal cancer. Nat Genet 2001;29:137–8.
21. Jiricny J, Hughes M, Corman N, Rudkin BB. A human 200-kDa protein binds selectively to DNA fragments containing G·T mismatches. Proc Natl Acad Sci U S A 1988;85:8860–4.
22. Ciotta C, Ceccotti S, Aquilina G, et al. Increased somatic recombination in methylation tolerant human cells with defective DNA mismatch repair. J Mol Biol 1998;276:705–19.
23. Risinger JI, Umar A, Barrett JC, Kunkel TA. A hPMS2 mutant cell line is defective in strand-specific mismatch repair. J Biol Chem 1995;270:18183–6.
24. DuBridge RB, Tang P, Hsia HC, Leong PM, Miller JH, Calos MP. Analysis of mutation in human cells by using an Epstein-Barr virus shuttle system. Mol Cell Biol 1987; 7:379–87.
25. Trojan J, Zeuzem S, Randolph A, et al. Functional analysis of hMLH1 variants and HNPCC-related mutations using a human expression system. Gastroenterology 2002;122:211–9.
26. Cejka P, Stojic L, Mojas N, et al. Methylation-induced G(2)/M arrest requires a full complement of the mismatch repair protein hMLH1. EMBO J 2003;22:2245–54.
27. di Pietro M, Marra G, Cejka P, et al. Mismatch repair-dependent transcriptome changes in human cells treated with the methylating agent MNNG. Cancer Res 2003;63:8158–66.
28. Marra G, Iaccarino I, Lettieri T, Roscilli G, Delmastro P, Jiricny J. Mismatch repair deficiency associated with overexpression of the MSH3 gene. Proc Natl Acad Sci U S A 1998;95:8568–73.
29. Thomas DC, Roberts JD, Kunkel TA. Heteroduplex repair in extracts of human HeLa cells. J Biol Chem 1991;266:3744–51.
30. Kondo E, Horii A, Fukushige S. The interacting domains of three MutL heterodimers in man: hMLH1 interacts with 36 homologous amino acid residues within hMLH3, hPMS1 and hPMS2. Nucleic Acids Res 2001;29:1695–702.
31. Raschle M, Marra G, Nystrom-Lahti M, Schar P, Jiricny J. Identification of hMutLβ, a heterodimer of hMLH1 and hPMS1. J Biol Chem 1999;274:32368–75.
32. Esteller M. Aberrant DNA methylation as a cancer-inducing mechanism. Annu Rev Pharmacol Toxicol 2005;45:629–56.
33. Liu L, Zhang J, Bates S, et al. A methylation profile of *in vitro* immortalized human cell lines. Int J Oncol 2005; 26:275–85.
34. Wang TF, Kung WM. Supercomplex formation between Mlh1-Mlh3 and Sgs1-Top3 heterocomplexes in meiotic yeast cells. Biochem Biophys Res Commun 2002; 296:949–53.
35. Chen PC, Dudley S, Hagen W, et al. Contributions by MutL homologues Mlh3 and Pms2 to DNA mismatch repair and tumor suppression in the mouse. Cancer Res 2005;65:8662–70.

7. Acknowledgements

I would like to thank Prof. Josef Jiricny for giving me the possibility to make my thesis at his institute, and for his excellent, and always stimulating, scientific discussions and advices.

I would like to specially thank Dr. Giancarlo Marra for the excellent supervision and guidance of my projects.

My sincere gratitude to all the institute members I have met over these years, especially to those that have contributed to the results presented in this thesis.

Special thanks to my lab mates Anne Anstett, Elda Cannavo, Elisa Catteneo, Ritva Haider and Emilija Veljkovic for creating a wonderful atmosphere and for their unconditional support whenever I needed.

I am glad to have shared fantastic moments and conversations with Stefano Ferrari, Franziska Fisher, Marianne Köpfler, Giancarlo Marra, Elisabetta Pani, Helga Pletscher, Kanagaraj Radhakrishnan, Malika Salah, Silvia Schanz, Anne Muller and, very especially, with Elda Cannavo, Elisa Cattaneo, Petr Cejka, Mahmoud El-Shemerly, Javier Peña Díaz, Emanuele Valtorta, and Emilija Veljkovic. You are all very nice persons.

Many thanks to former and present technical assistants for their efficient effort which enormously facilitates our research work: Helga Pletscher, Marianne Grima, Marianne Köpfler, Jana Rojickova, Patrick Greiner, Margaret Fäsi, Ritva Haider, Christine Hemmerle, Christiane König, Ester Kohler, Christoph Moser, Malika Salah, Najat Maanaoui-Salah, Farah M'Hamedi-Baccouche, and Peter Binz.

I would like to emphasize the great assistance received from Marzana Künzli-Gotarczyk, Endre Laczko, Andrea Patrignani, Hubert Rehrauer and Ulrich Wagner at the FGCZ. Thank you very much for your help and patience.

I would like to thank and recognize the awesome contribution to this work done by all our collaborators, especially Mariagrazia de Palo and Laurens Van der Flier. It is a pleasure to work with you.

I would also like to thank Alex Gil-Vernet for the great time we had together in Zurich. I will always remember how much fun we had in "La Penya" watching the Barça games!!!... We have to start with it again when we definitely come back to Barcelona.

Thanks to Ignacio Fernández de las Nieves (Nachorrrrrrl) for the great moments and lots of fun we had together (the "Squash Battles" were memorable!!!). To discuss with you was always a pleasure and my little knowledge about Organic Chemistry is thanks to you… jejejjejejeje!!!!.

Thanks to Patricia Fletcher for her help and patience with my English, I really appreciate it.

Many many thanks to those FRIENDS who have kept in touch, despite the distance, and/or have come to visit us to Zürich: Marc Oliveras, Albert Ramón & Mar, Carolina Mordillo, Cristina Vilella & Borja, Alex Gil-Vernet & Anna, Fadhel Ben Chaabane, Reinaldo Fraga, Albert Nebra, Ernest Millán, Xavier Rubiralta, Alberto Román, Nestor Busquets, and Alfonso Chico de Guzmán. It has been really nice from you.

Thanks to my family for their constant and total support that helped me to get so far. Thanks to have made me feel so close being so far. Thanks to: i) my mother, Roser, for her curiosity about my

research work and for her interesting critical point of view, ii) my father, Ramón, to make little things unforgettable, like a breakfast at Mövenpick, with his enthusiasm, iii) my brother, Aaron, and his wonderful family; Marta, Rita and Ramón Jr. because I feel extremely happy everytime I see them, iv) my sister, Raquel, and Joan for having me punctually informed about what is going on in Barcelona via "Radio Macuto", v) my sister, Carla, and Carles to have created such a nice creature: Vico, and vi) Nestor (yes, you are part of the family… jejejjejeej!!!) because our friendship is so strong that I am sure it will last forever. Of course, thanks to all of you to have made it to come to my defense, it was very important for me.

Finally, thanks to Carlota. Gràcies per ser com ets (dolça, divertida, intel.ligent, tenaç, emprenedora, guapa, carinyosa, amb caràcter... en definitiva… meravellosa!!!) i per estimar-me com m'estimes. Compartir la meva vida amb tu es el millor que em podia passar. Em fas sentir la persona més afortunada d'aquest món… Tatimu.

8. *Curriculum Vitae*

Personal Information

First name	Jacob
Last name	Sabatés Bellver
Date of birth	7th October 1978
Place of birth	Barcelona, Spain
Nationality	Spanish

Education

1984-1998	High school: Institut Tècnic Eulàlia, Barcelona, Spain
1991	High school: Vicente Cañada Blanch, London, UK
1998-2003	University: **Licensed in Biotechnology** (10 semesters), Universitat Autònoma de Barcelona, Barcelona, Spain
2003	Diploma thesis: **"Bioalcohol production by S.cerevisiae at high cellular densities in a two-stage bioreactor with recirculation of biomass through membranes"** (1 semester), Department de Génie Biochemique et Alimentaire (GBA), Laboratoire Biotechnologie-Bioprocédés (LBB), Institute National des Sciences Apliquées (INSA), Toulouse, France. Supervised by Dr.Carole Molina-Jouve and Prof. Dr. Gérad Goma
2003-2006	PhD studies: **"The Transcriptome of Epithelial Transformation in the Distal Gastrointestinal Tract"** (7 semesters), Institute of Molecular Cancer Research of Zürich (IMCR), University of Zürich, Zürich, Switzerland. Supervised by Dr. Giancarlo Marra and Prof. Dr. Josef Jiricny

9. List of publications

di Pietro M*, **Sabates Bellver J***, Menigatti M, Bannwart F, Schnider A, Russell A, Truninger K, Jiricny J, Marra G. “Defective DNA Mismatch Repair Determines a Characteristic Transcriptional Profile in Proximal Colon Cancers” *Gastroenterology* 2005; **129**:1047-1059. *These authors contribute equally to this work.

Cannavo E, Marra G, **Sabates-Bellver J**, Menigatti M, Lipkin SM, Fisher F, Cejka P, Jiricny J. “Expression of the MutL Homologue hMLH3 in Human Cells and its Role in DNA Mismatch Repair” *Cancer Res* 2005; **65** (23): 10759-66.

Murtaza I, Marra M, Schlapbach R, Patrignani A, Künzli M, **Sabates J**, Dutt A. “A preliminary investigation demonstrating effect of quercetin on expression of genes related to cell cycle arrest, apoptosis and xenobiotic metabolism in human CO115 colon adenocarcinoma cells using DNA microarray” *Biotechnol Appl Biochem* 2006; **45**:29-36.

Van der Flier LG*, **Sabates-Bellver J***, Oving I, Haegebarth A, de Palo M, Anti M, van Gijn ME, Suijkerbuijk S, Van de Wetering M, Marra M, Clevers H. “The intestinal Wnt/TCF signature” *Gastroenterology 2007; **132**:628-32*. *These authors contribute equally to this work.

Sabates-Bellver J*, Van der Flier LG*, de Palo M, Cattaneo E, Maake C, Rehrauer H, Laczko E, Kurowski MA, Bujnicki JM, Menigatti M, Luz J, Ranalli TV, Pastorelli A, Faggiani R, Anti A, Jiricny J, Clevers H, Marra G. “Transcriptome profile of human colorectal adenomas and a novel biomarker of epithelial transformation” In press, Molecular Cancer Research. *These authors contribute equally to this work.

Sabates-Bellver J, Gonzenbach T, Rehrauer H, Laczko E, Maake C, Bannwart F, Schnider A, Jiricny J, Marra G. “The transcriptome of ileal and colonic normal mucosa” Manuscript in preparation.

10. Abstracts presented in scientific meetings

*"**Gene Expression Profile in Colorectal Cancer**"*, RRR Club (DNA Replication, Repair and Recombination Club), 23th March 2004, IMCR, Zürich, Switzerland. Jacob Sabatés Bellver (**short talk**).

*"**The transcriptional profile of colon adenomas differs dramatically from that of the normal colonic mucosa**",* XI National Congress of the Italian Federation of Digestive Diseases, 12-16 March 2005, Geneva, Italy. Jacob Sabates-Bellver, Mariagrazia de Palo, Mirco Menigatti, Alfredo Pastorelli, Roberto Faggiani, Giovanna Verrico, Vito Gomes, Andrea Patrignani, Ulrich Wagner, Luca Roncucci, Maurizio Ponz de Leon, Marcello Anti, Josef Jiricny and Giancarlo Marra (**poster, awarded**, presented by M. de Palo).

*"**Gene Expression Profile in Colorectal Adenomas**"*, VII Charles Rodolph Brupbacher Scientifyc Symposium, 16-19 March 2005, Zurich, Switzerland. Jacob Sabates-Bellver, Mariagrazia de Palo, Mirco Menigatti, Alfredo Pastorelli, Roberto Faggiani, Giovanna Verrico, Vito Gomes, Andrea Patrignani, Ulrich Wagner, Luca Roncucci, Maurizio Ponz de Leon, Marcello Anti, Josef Jiricny and Giancarlo Marra (**poster**).

"Difference between the transcriptional profile of the adenomatous mucosa and the normal mucosa of the colon", 12th National Congress of Digestive Diseases, 1-5 April 2006, Naples, Italy. Sabates-Bellver J., de Palo M., Menigatti M., Gomes V., Pastorelli A., Faggiani R., Caturelli E., Zampaletta C., Jiricny J., Marra G., Anti M. (**poster** presented by M. de Palo).

"Identification of the genetic signature of the colon cancer characterized by microsatellite instability by microarray analysis", 12th National Congress of Digestive Diseases, 1-5 April 2006, Naples, Italy. di Pietro M.*, Sabates-Bellver J.*, Menigatti M., Bannwart F., Schnider A., Truninger K., Jiricny J., Marra G. (**poster** presented by M. di Pietro). *These authors contributed equally to this work.

"Expression profile of Wnt signalling genes in human normal ileal and colon mucosa", Wnt and beta-Catenin Signaling in Development and Disease Congress, 7-12 April 2006, Snowbird, Salt Lake City, UT, USA. Jacob Sabates-Bellver, Tobias Gonzenbach, Fridolin Banwart, Annalies Schnider, Josef Jiricny and Giancarlo Marra (**poster**).

"The Wnt/TCF signature", Wnt and beta-Catenin Signaling in Development and Disease Congress, 7-12 April 2006, Snowbird, Salt Lake City, UT, USA. Laurens G. Van der Flier*, Jacob Sabates-Bellver*, Irma Oving, Andrea Haegebarth, Mariagrazia de Palo, Marcello Anti, Marielle E. van Gijn, Saskia Suijkerbuijk, Marc Van de Wetering, Giancarlo Marra and Hans Clevers (**poster** presented by L.G. Van der Flier). *These authors contributed equally to this work.

"KIAA1199: A Putative Wnt Signaling Target Gene Over-expressed in Colorectal Tumors", 2nd Cancer Research Retreat, 1-3 September 2006, Monte Verita, Ascona, Switzerland. Jacob Sabates-Bellver*, Laurens G. Van der Flier*, Mariagrazia de Palo, Marcello Anti, Marc Van de Wetering, Josef Jiricny, Hans Clevers and Giancarlo Marra (**short talk**). *These authors contributed equally to this work.

"The transcriptome of human colorectal adenomas: Identification of KIAA1199 as a novel tumor biomarker", 6th ISREC Conference on Cancer Research, 11-13 October 2006, Lausanne, Switzerland. Jacob Sabates-Bellver*, Laurens G. Van der Flier*, Mariagrazia de Palo, Elisa Cattaneo, Caroline Maake, Hubert Rehrauer, Endre Laczko, Michal A. Kurowski, Janusz M. Bujnicki, Mirco Menigatti, Judith Luz, Teresa V. Ranalli, Alfredo Pastorelli, Roberto Faggiani, Marcello Anti, Josef Jiricny, Hans Clevers, and Giancarlo Marra (**poster**). *These authors contributed equally to this work.

"Colorectal Adenomas and Normal Mucosa: Differences between Transcriptional Gene Expression Profile", 14th United European Gastroenterology Week, 21-25 October 2006, Berlin, Germany. *Gut* supplement n° V, Vol 55 – *Endoscopy* supplement n° II, Vol 38, pag. A69, OP-G-295. Sabates-Bellver J., de Palo M., Menigatti M., Gomes V., Ranalli T., Pastorelli A., Faggiani R., Caturelli E., Zampaleta C., Jiricny J., Marra G., Anti M. (**short talk** given by M. de Palo).

"The Transcriptome of Colorectal Adenomas", RRR Club (DNA Replication, Repair and Recombination Club), 19th April 2007, IMCR, Zürich, Switzerland. Jacob Sabatés Bellver (**short talk**).